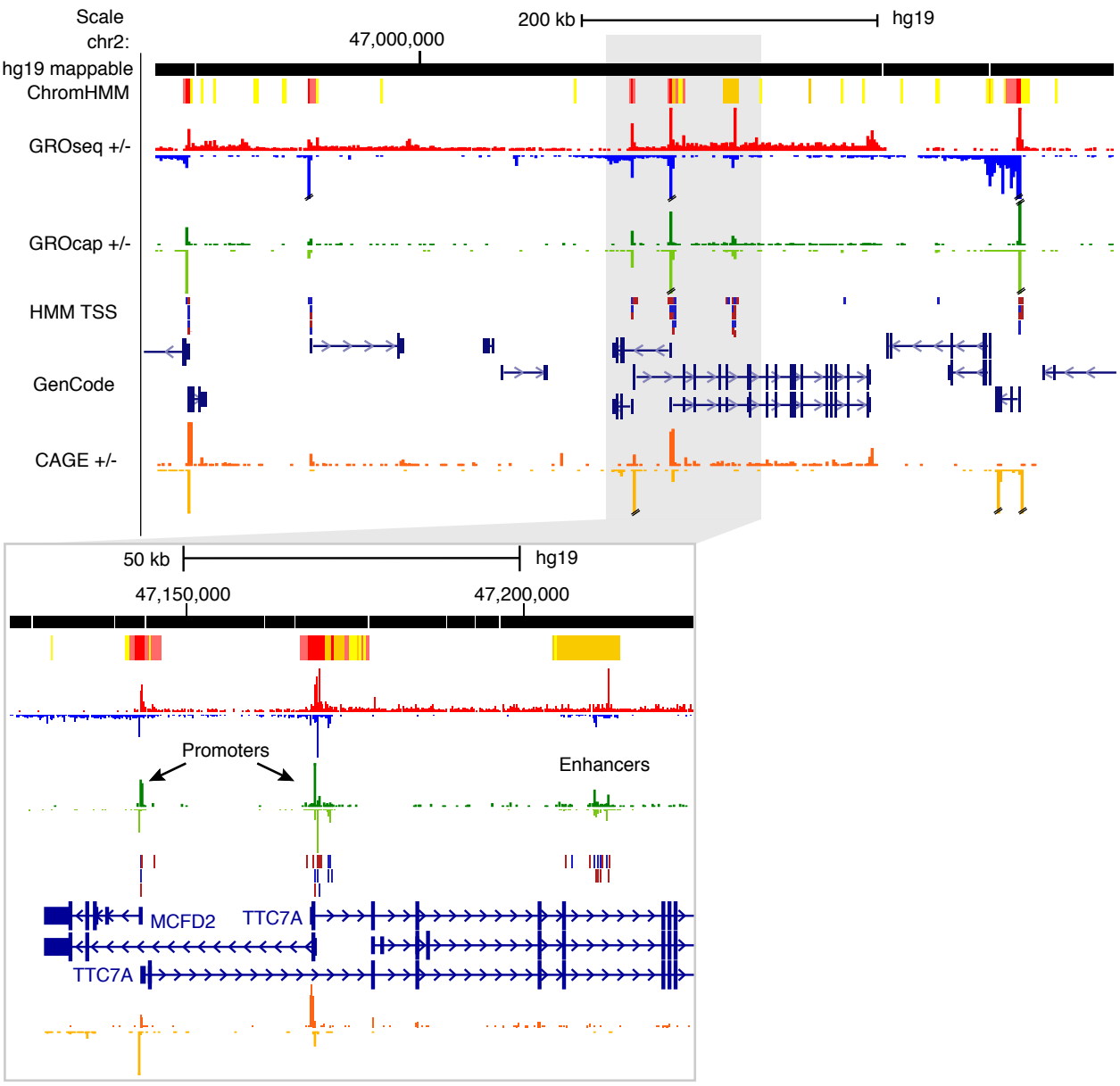
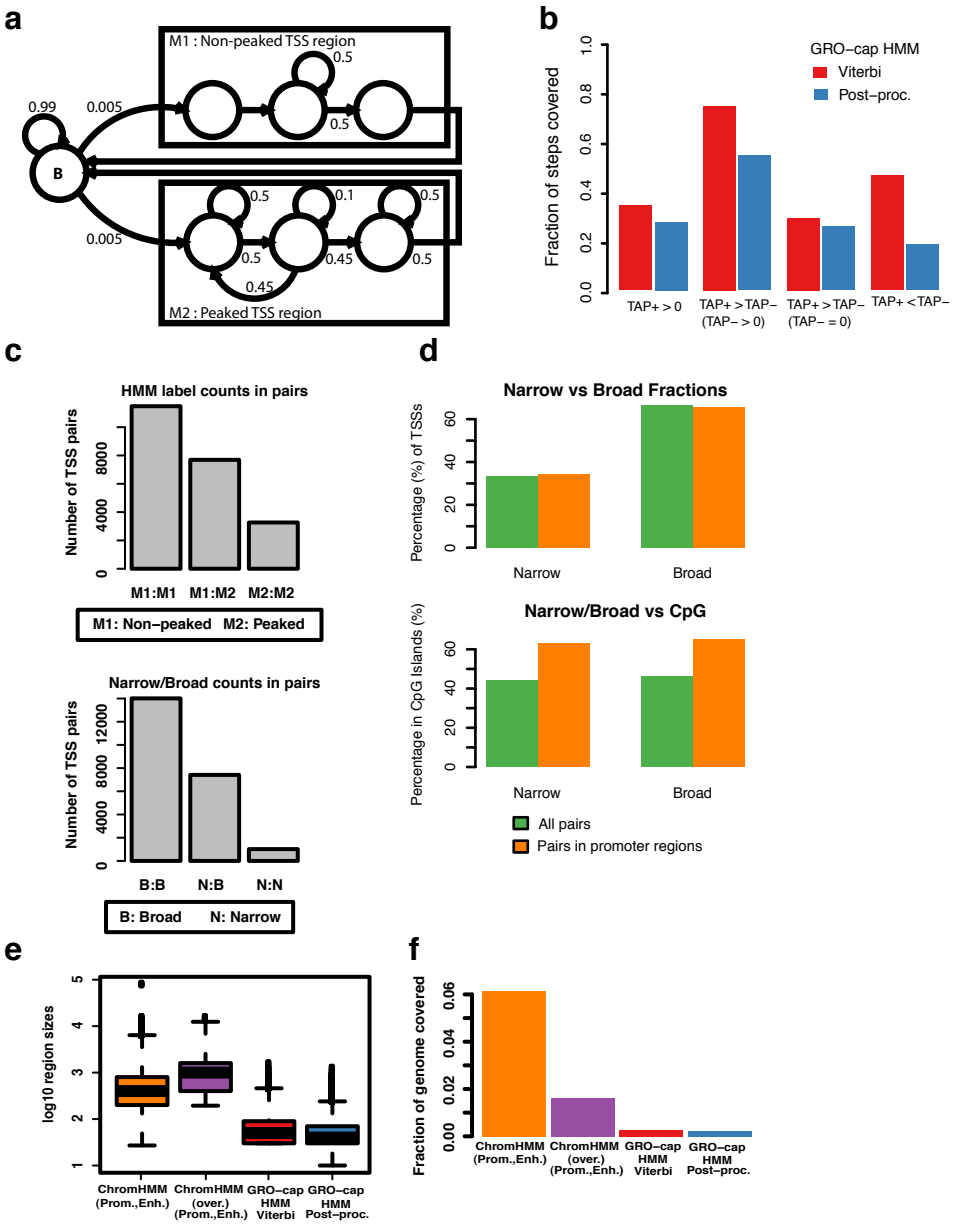
Core and Martins et al. Supplemental Figure 1

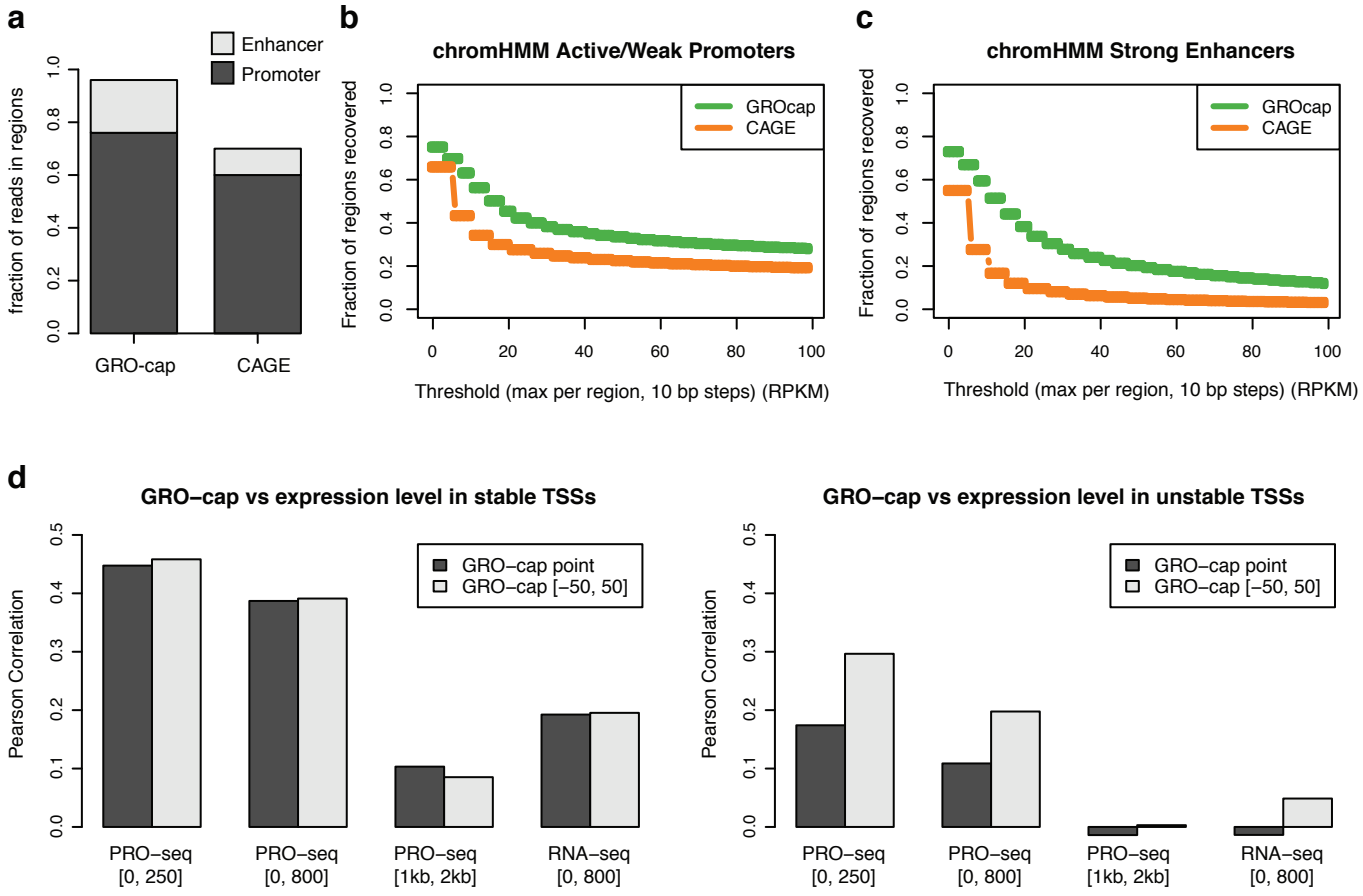


Supplemental Figure 1. Browser shot of genomic data used in this study
Shows a browser shot from the UCSC genome browser showing some of the data sets generated in this study (or previously published). The inset is a zoomed in view of the shaded region that shows the divergent GRO-cap (+ strand: dark green, - strand: light green) signal at a couple promoters (ChromHMM28: red) and enhancers (ChromHMM: orange). Note that CAGE signal (+ strand: dark orange, - strand: light orange) is at background levels in the enhancer region. Data is from GM12878 cells.



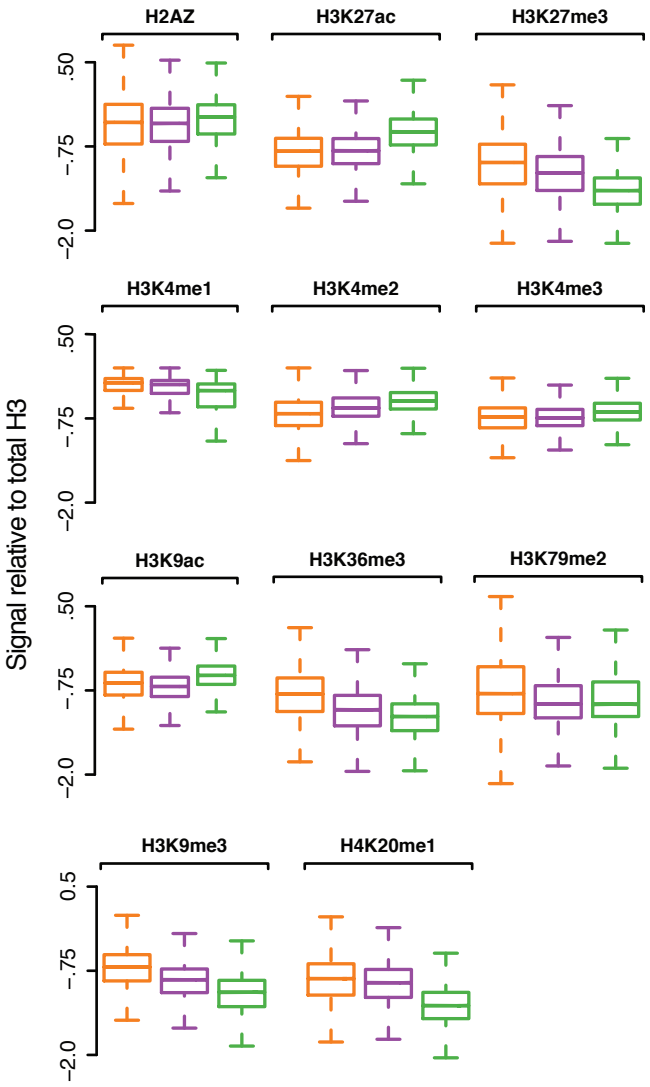
Supplemental Figure 2. TSS Identification from GRO-cap

(a) Hidden Markov model state diagram, with three state groups: B) background; M1) non peaked TSS region; M2) peaked TSS region (non-trivial transition probabilities indicated as labels to arrows). **(b)** Effect of TSS region prediction post-processing (see Methods) on coverage of GRO-cap data split by relationship between pre and post TAP signal. Overall, there is a small reduction in the fraction of the GRO-cap library that is covered (in number of 10 bp steps), with the largest reduction falling on depleted steps (TAP+ < TAP-). **(c)** Number of TSS pairs that correspond to each combination of peaked (M2) and non-peaked (M1) subsets (top) and each combination of broad (B) and narrow (N). **(d)** Narrow/broad distinction based on whether over less/more than 50% of GROcap reads are within +/- 2bp of the mode (best site). There is 45% agreement between the two ways to label pairs (assuming M1 = B and M2 = N). Comparison of narrow and broad TSS regions (from paired subset) with promoter annotations (ChromHMM, top panel) and CpG Island overlap (bottom panel; CpG Island track from the UCSC Genome Browser). Narrow/broad distinction based on whether over less/more than 50% of GROcap reads are within +/- 2bp of the mode (best site). No significant difference is observed in either case. **(e)** Distribution of ChromHMM Promoter and Enhancer regions and GRO-cap TSS predictions lengths. For ChromHMM, we show both the full set (orange) and the subset that has an overlapping GRO-cap TSS prediction (purple). GRO-cap TSS prediction lengths are shown for both before (Viterbi, red) and after post-processing (blue) (see Methods). **(f)** Fraction of the genome covered by predicted TSS regions compared with ChromHMM Promoter and Enhancer regions. For ChromHMM, we show both the full set (orange) and the subset that has an overlapping GRO-cap TSS prediction (purple). GRO-cap TSS prediction lengths are shown for both before (Viterbi, red) and after post-processing (blue). Data are from GM12878 cells.



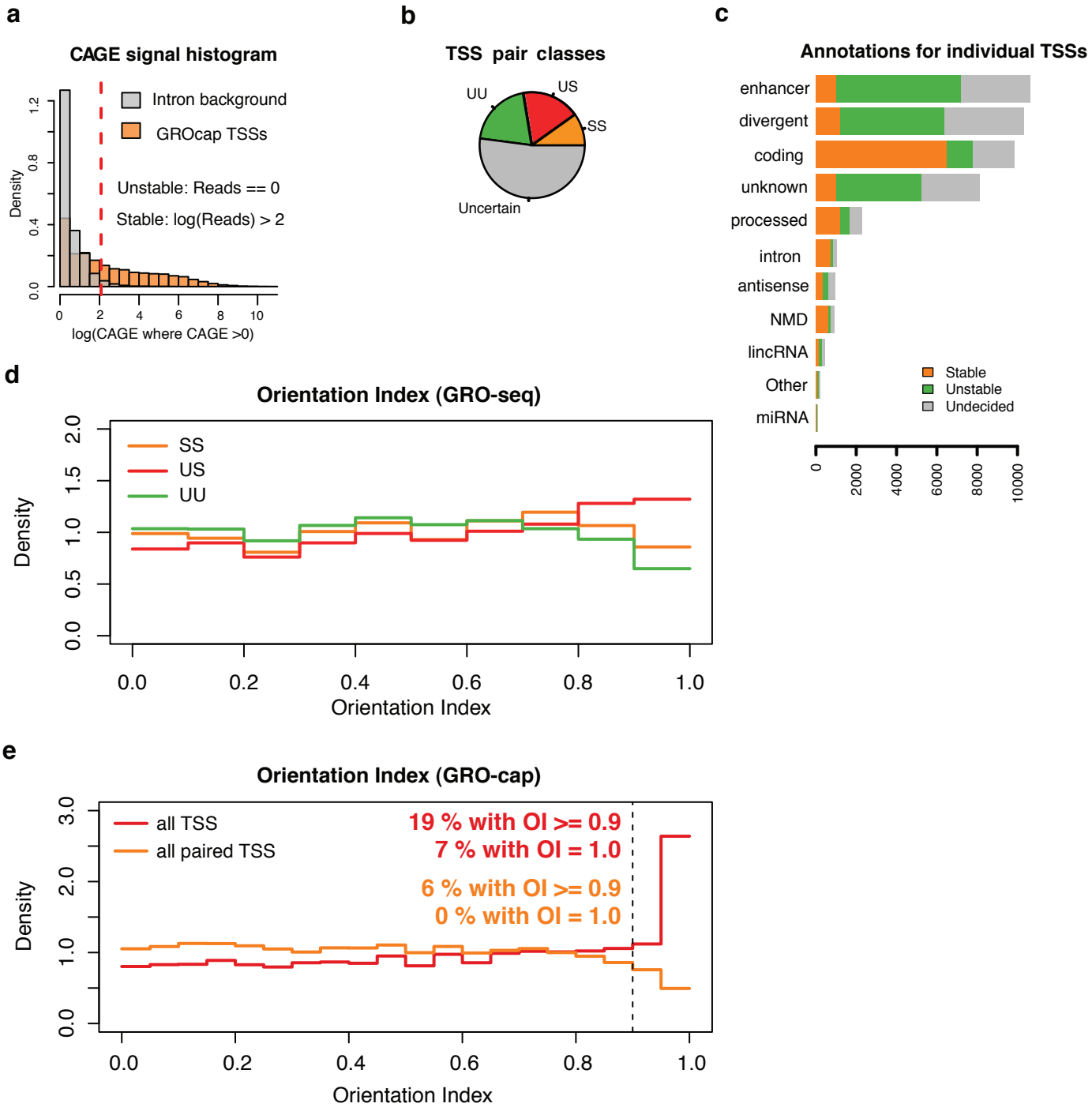
Supplemental Figure 3. Comparison of GRO-cap and CAGE

(a) Fraction of reads in promoters (dark grey) and enhancers (light grey) for GRO-cap (76% promoter, 20% enhancer) and CAGE (60% promoter, 10% enhancer). **(b,c)** Recovery threshold plots showing the fraction of total promoters **(b)**, and enhancers **(c)** that are recovered at varying thresholds of GRO-cap (green) and CAGE (orange). Data are from GM12878 cells. **(d)** Comparison of the correlation of GRO-cap data at the max TSS location (dark shade) or including the surrounding region (light shade) with PRO-seq or RNA-seq data at stable (left panel) and unstable (right panel) TSS pairs. Position noted underneath the bar graphs represent the positions relative to the TSS where the signals where tabulated for PRO-seq or RNA-seq.



Supplemental Figure 4. Histone modifications in enhancer classes
Distribution of ChIP-seq histone modification signals in each enhancer class (closed: orange, open: purple, transcribed: green), scaled by total H3K4 methylation signal.

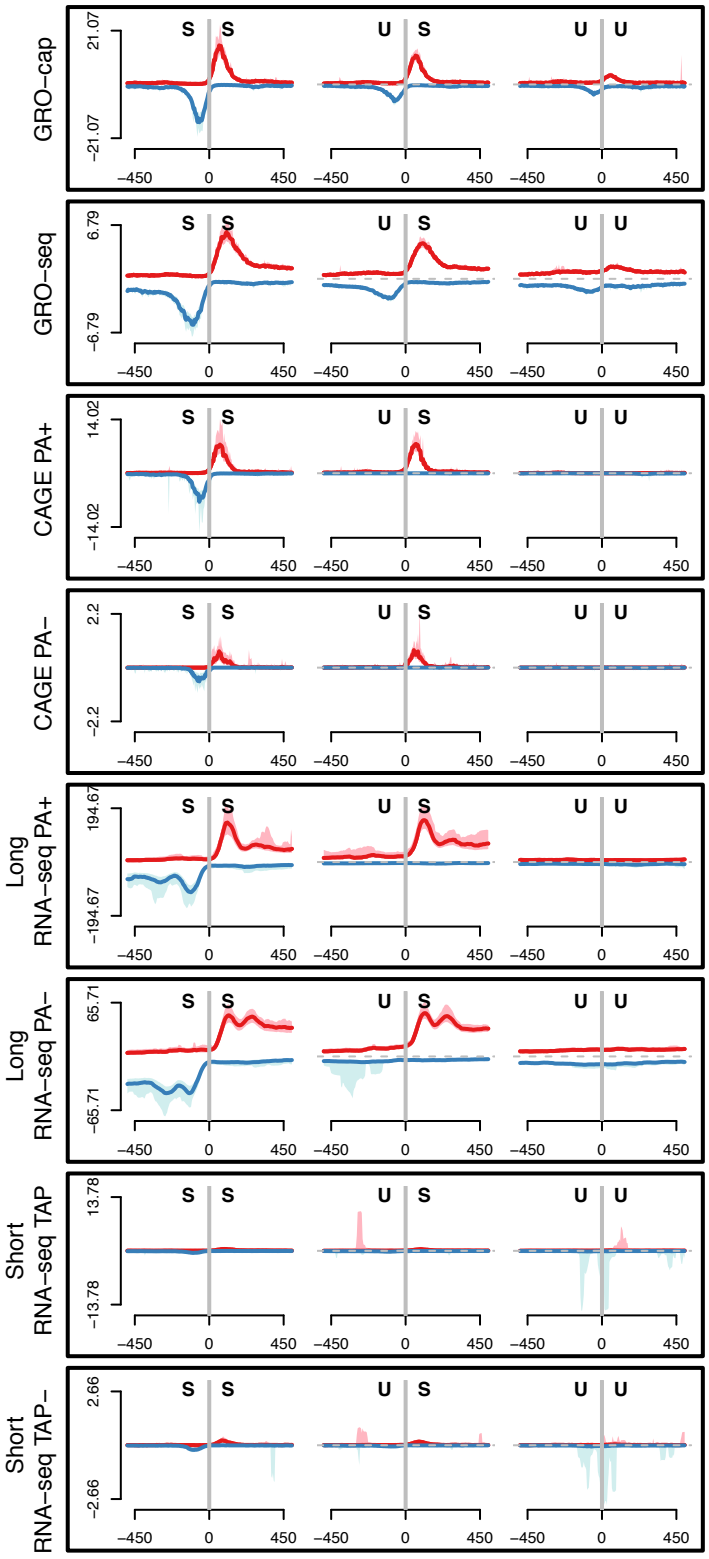
Core and Martins et al. Supplemental Figure 5



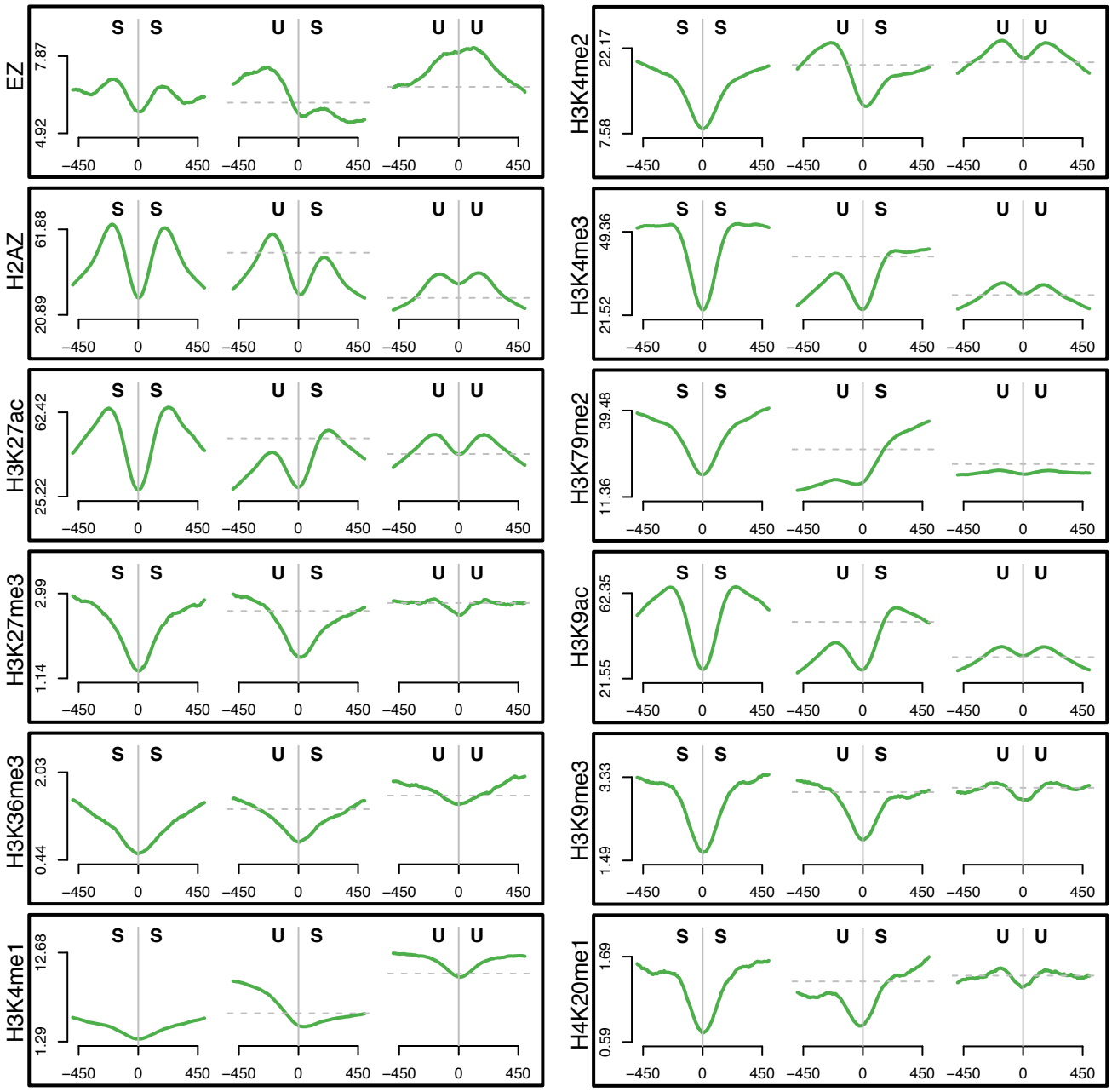
Supplemental Figure 5. Identification of TSS pair classes

(a) CAGE signal histogram at GRO-cap TSSs (orange) overlaid with CAGE background signal estimated from introns (grey). TSSs were classified as stable if above threshold indicated by dashed red line, or unstable if they contain no CAGE reads. **(b)** Pie chart shows relative proportion of TSS pair stability classes, including "Uncertain" for those in between the two thresholds. **(c)** Individual TSSs within pairs were matched to various annotations based on GENCODE annotations or ChromHMM regions (for enhancers). TSSs for each annotation were the split on stability classifications: stable (orange), unstable (green), undecided (gray). **(d)** Orientation indexes (OI) are presented for pairs classifies as stable::stable (red), unstable::stable (orange), unstable::unstable (green). OI scales between zero (bi-directional) and one (uni-directional) and is defined as 2×(max(Rp,Rm)/(Rp + Rm) – 1. Rp and Rm are the, plus and minus strand, respectively, GRO-seq reads that fall in the 250 bp downstream of the strongest (highest read count) GRO-cap position in the TSS region in each pair. OI was calculated from GRO-seq data in GM12878 cells. **(e)** Distribution of OI's calculated at all GRO-cap TSSs and after selection for GRO-cap TSS pairs. Aside from the small fraction of non-paired TSSs (7%), the distribution of OI is similar suggesting that the selected TSS pairs used inour analyses are a suitable representative set of TSSs.

Core and Martins et al. Supplemetal Figure 6

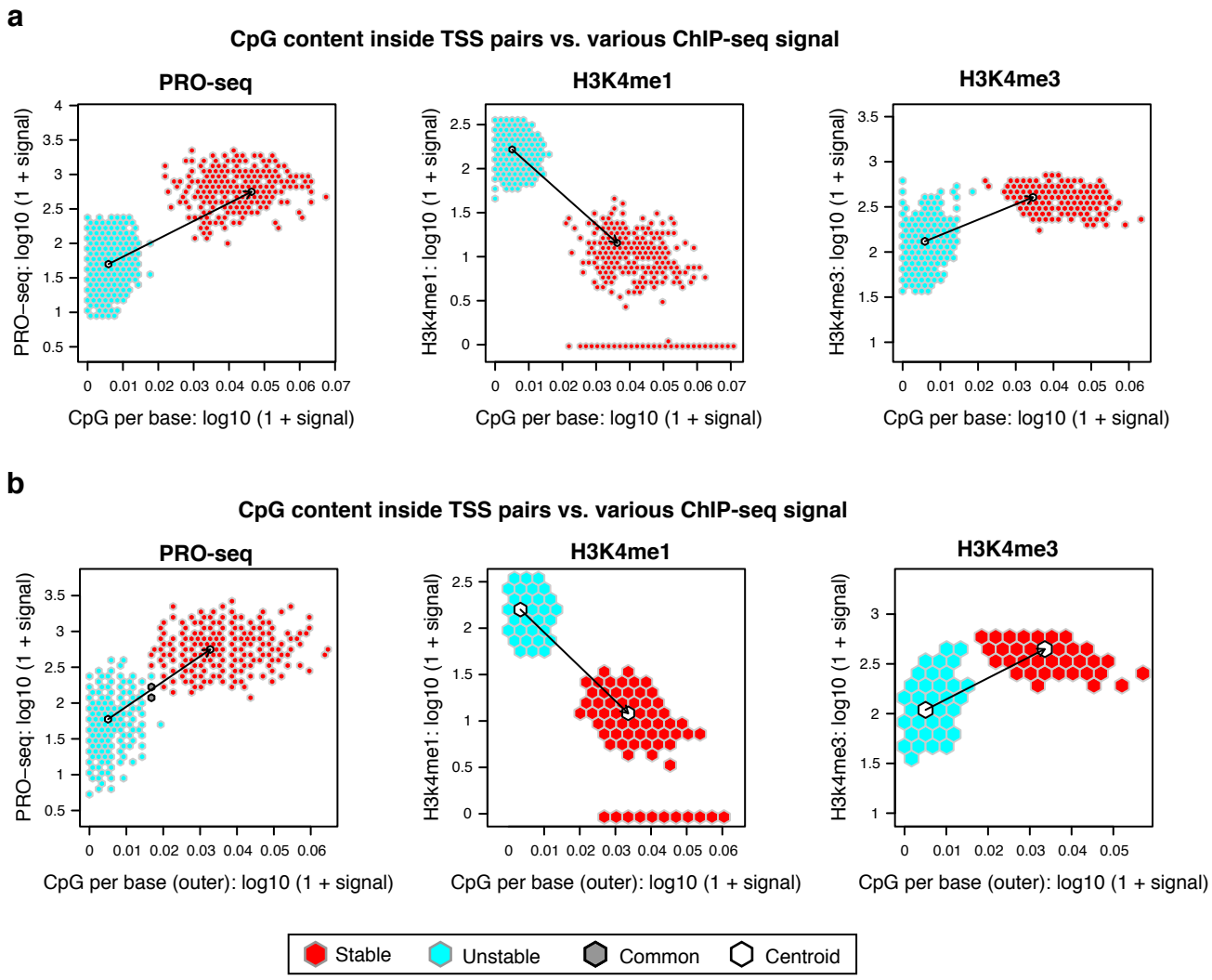


Supplemental Figure 6. Profiles of various RNA sequencing data at TSS pairs after stability classification
Metaplot profiles of various types of RNA sequencing data aligned to the center of GRO-cap TSS pairs after classifying pairs based on the stability of the transcript produced. Profiles are stable::stable (left), unstable::stable (center), unstable::unstable (right). GRO-seq and GRO-cap data were produced for this study. All other data were produced by the ENCODE consortium. Data are from GM12878 cells.



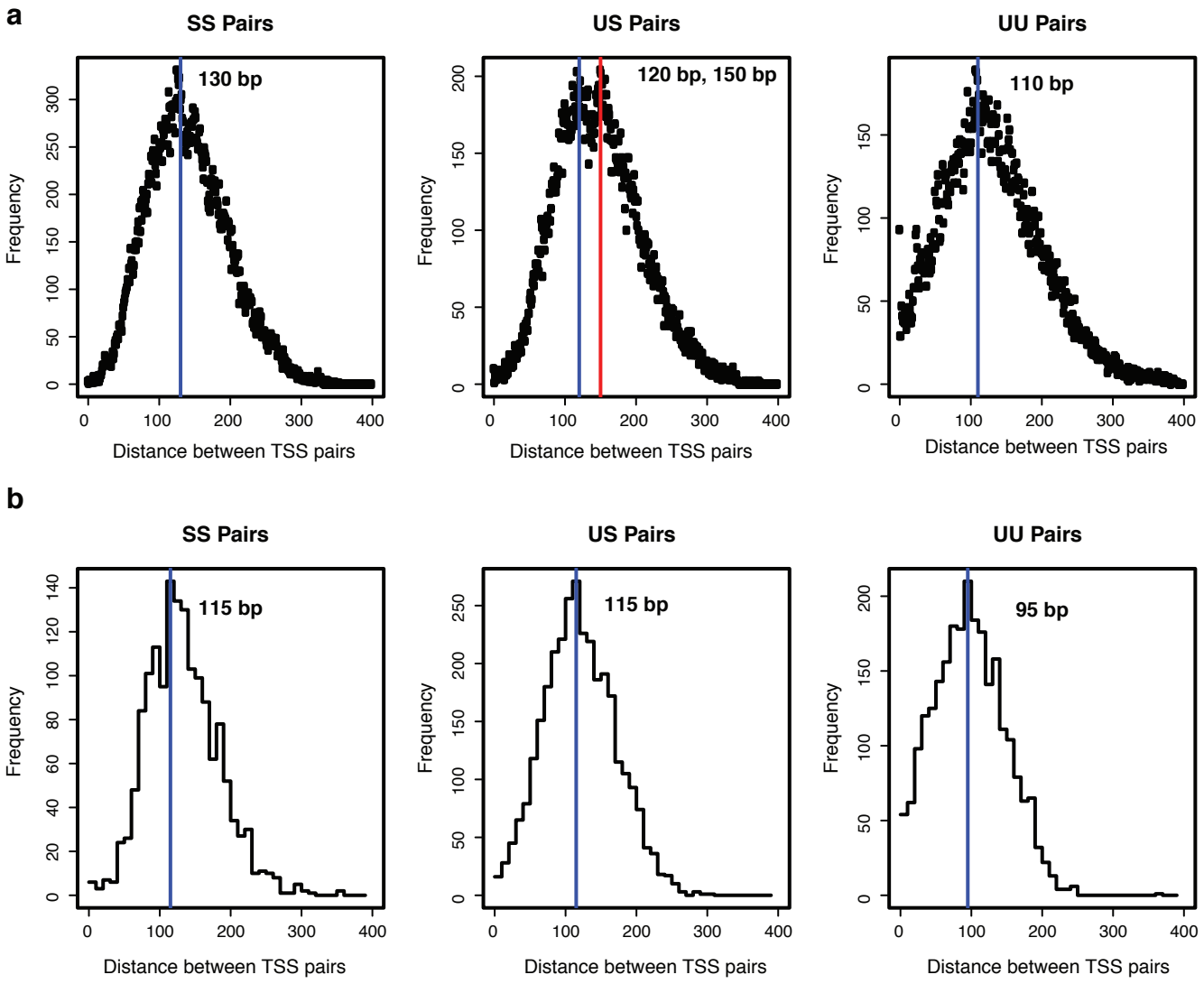
Supplemental Figure 7. Profiles of histone marks or chromatin binders at TSS pairs after stability classification
Composite profiles of ChIP-seq data for various histone modifications, variants, or chromatin binding proteins aligned to the center of GRO-cap TSS pairs after classifying pairs based on the stability of the transcript produced. Profiles are stable::stable (left), unstable::stable (center), unstable::unstable (right). All ChIP-seq data were produced by the ENCODE consortium. GRO-cap data were from GM12878 cells.

Core and Martins et al. Supplemental Figure 8



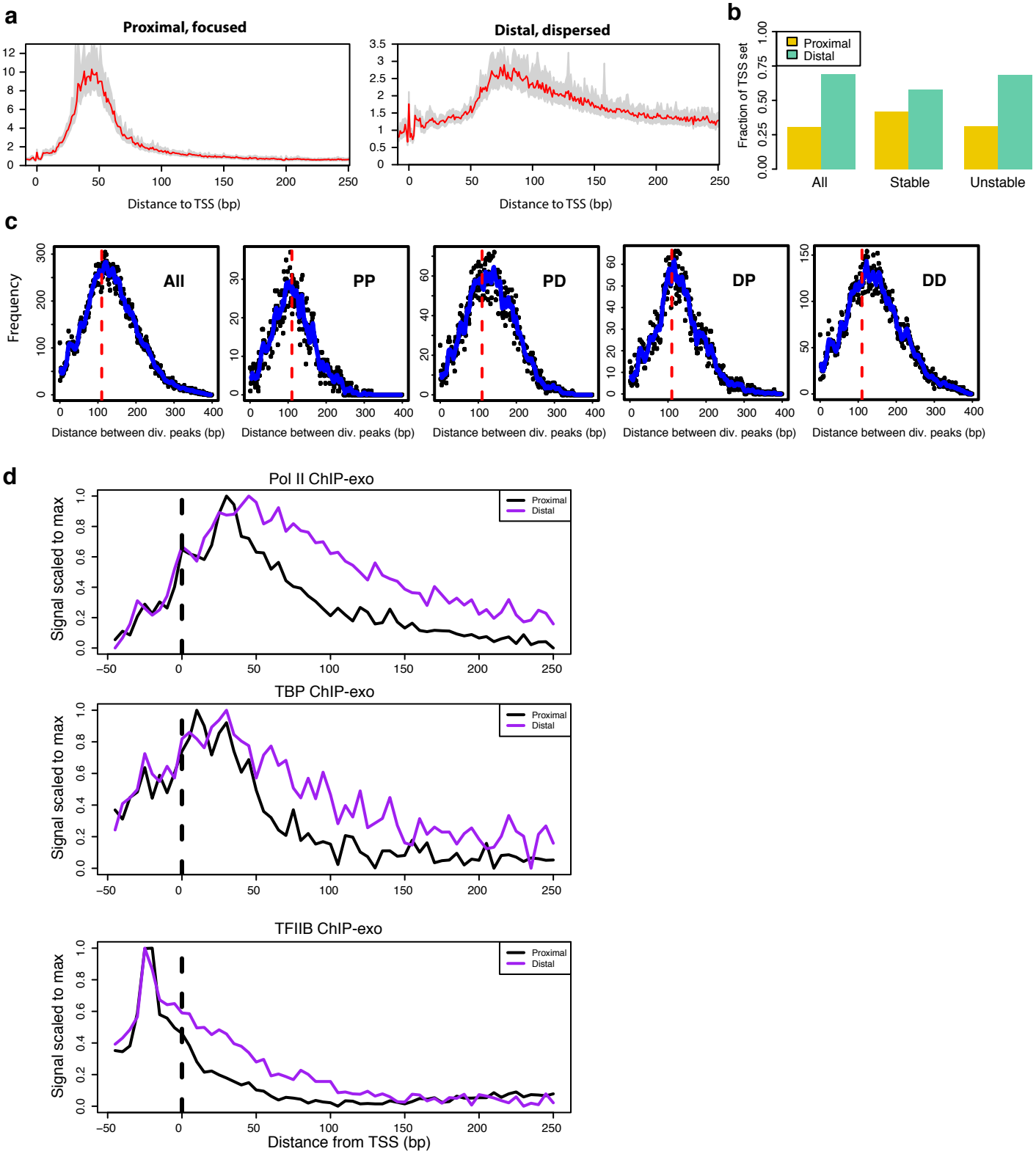
Supplemental Figure 8. CpG content vs. transcription and histone modifications at divergent TSSs
(a) CpG content inside TSS pairs versus PRO-seq signal (left), H3K4me1 (center) and H3K4me3 (right). Signal is further split between unstable (light blue) and stable (red) TSSs. Centroid for each subset in white. **(b)** CpG content outside TSS pairs versus PRO-seq signal (left), H3k4me1 (center) and H3k4me3 (right). Signal is further split between unstable (light blue) and stable (red) TSSs. Centroid for each subset is in white. PRO-seq data are from K562 cells.

Core and Martins et al. Supplemental Figure 9



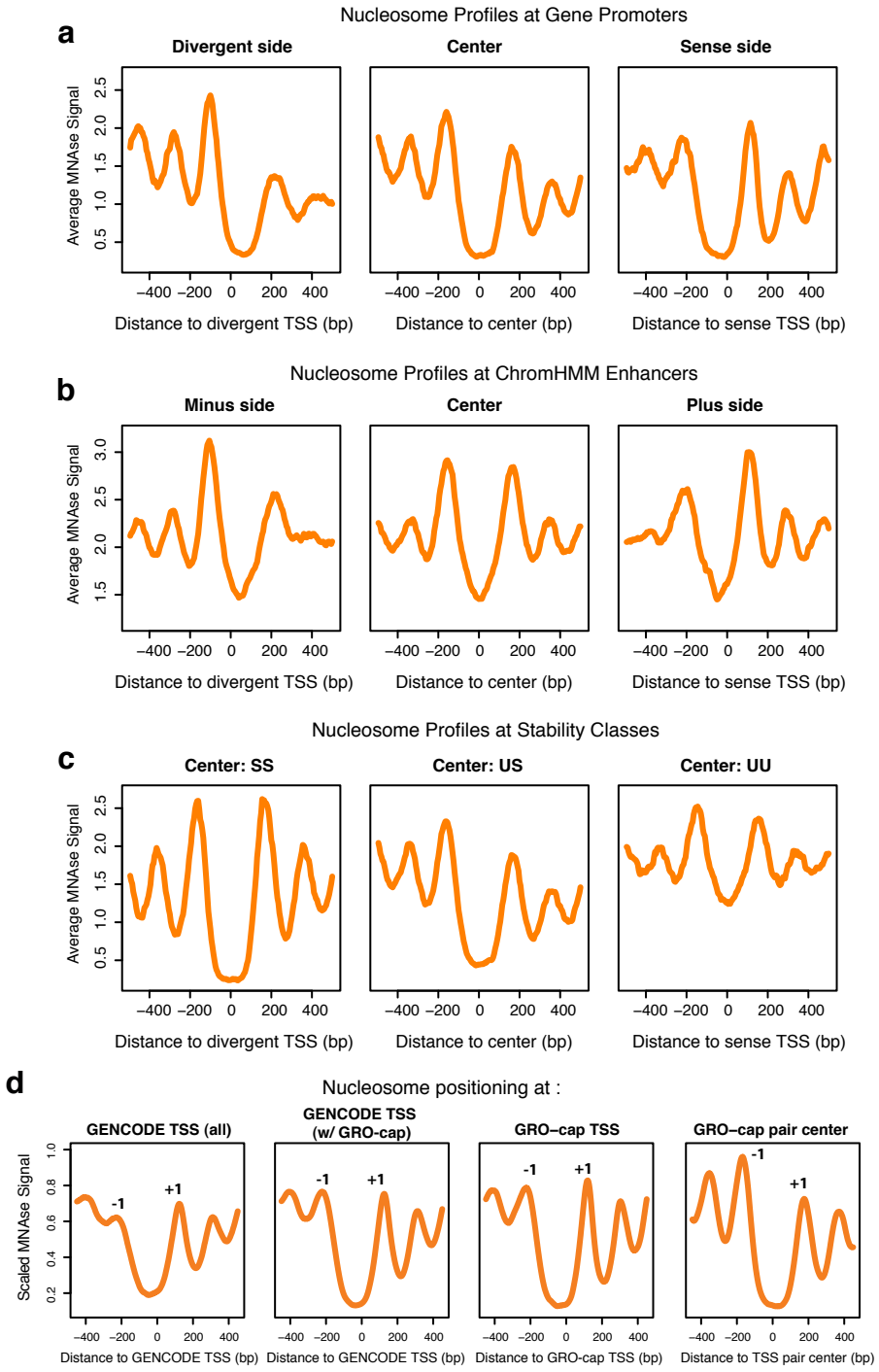
Supplemental Figure 9 Estimates of Inter-pair TSS distances at TSS with different stability classifications
(a) Divergent TSS distance distribution obtained by computing, over all TSS pairs, the distances between each combination of plus strand and minus strand positions within +/- 150bp of the divergent TSS center and where the GRO-cap signal is significantly above the control signal. Separate estimates obtained for each stability class (SS, US and UU). Estimates vary between 110 and 150 bp. **(b)** Divergent TSS distance distribution obtained by computing, over all TSS pairs, the distance between the centers of mass of the TSS regions in each strand. Estimates vary between 95 and 115 bp.

Core and Martins et al. Supplemental Figure10



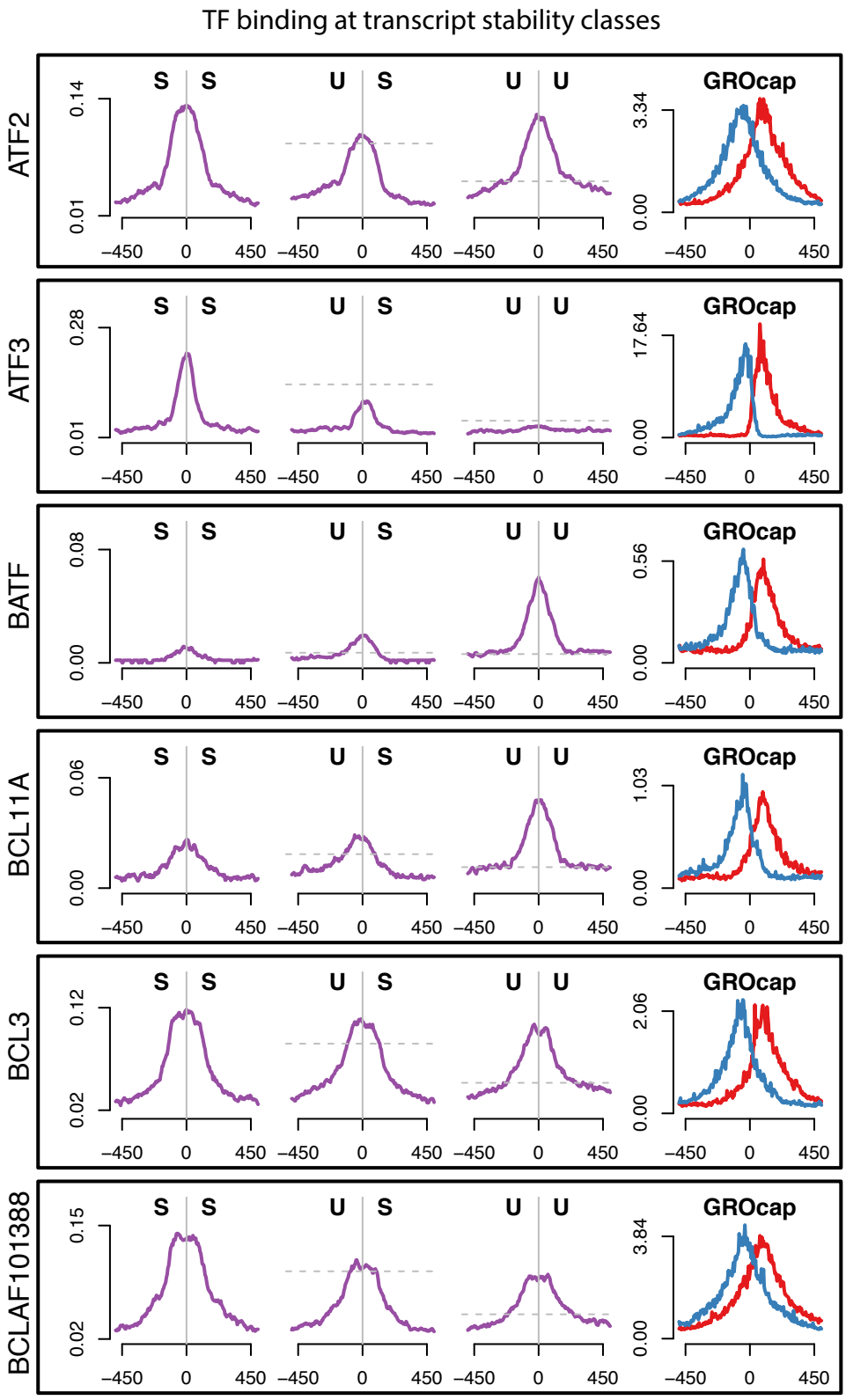
Supplemental Figure 10. Promoter-proximal pause versus TSS distances in pairs
(a) Two modes of promoter-proximal pausing are detectable (via k-means clustering): proximal-focused and distal-dispersed; this is consistent with previous results in Drosophila. **(b)** Comparison of promoter-proximal pause modes with TSS stability classes shows an enrichment of distal-dispersed pause mode in unstable versus stable and an overall preference for distal-dispersed pausing across all TSSs. **(c)** TSS distances between divergent TSSs (in pairs) segregated by pause mode labeling on each side (P for proximal-focused, D for distal-dispersed). There is no apparent effect of pausing mode on distance. **(d)** ChIP-exo data composite plots of Pol II (top), TBP (middle) and TFIIB (bottom) aligned to GRO-cap TSS at both proximal-focused and distal-dispersed pause mode subsets. Note that ChIP-exo does not necessarily represent the position of each factor as they can cross-link to the DNA through other factors.
Data are from K562 cells.

Core and Martins et al. Supplemental Figure 11

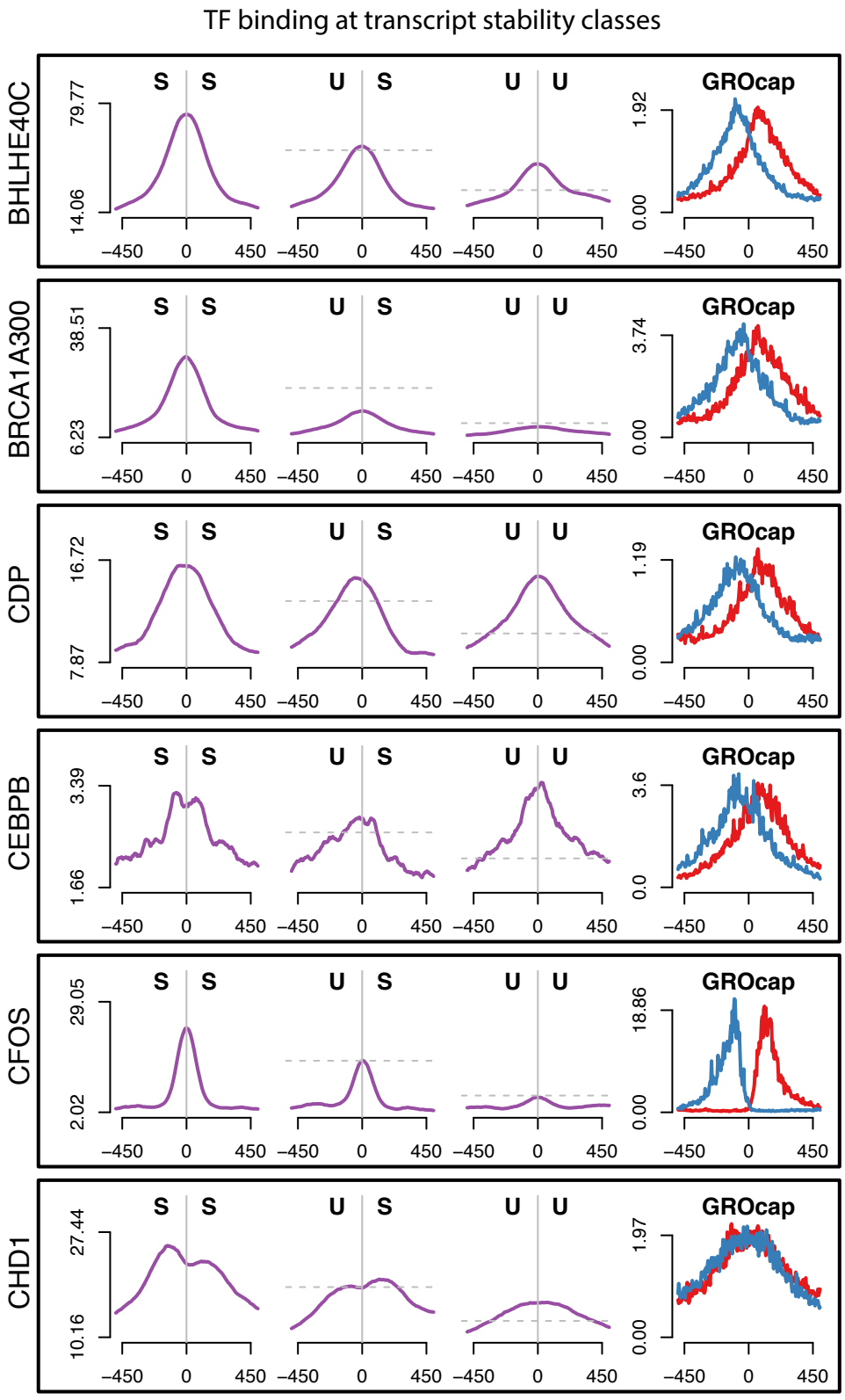


Supplemental Figure 11. Nucleosome profiles at TSS pairs
Nucleosome profiles at TSS pairs that map to **(a)** promoters and **(b)** enhancers. MNase-seq data is aligned to upstream TSS (left, divergent), the center of the pairs (center) or the downstream TSS within the pairs (right, sense). **(c)** Shows the nucleosome profiles aligned to the center of pairs after classifying pairs based on the stability of the transcript produced. Profiles are stable::stable (left), unstable::stable (center), unstable::unstable (right). MNase-seq data and GRO-cap data are from GM12878 cells. **(d)** Nucleosome profiles at all GENCODE TSSs (left panel), GENCODE TSSs that also have corresponding GRO-cap signal (left, middle panel), GRO-cap TSSs encoding the gene (middle right panel), and the center of GRO-cap TSS pairs. The -1 nucleosome becomes progressively more positioned as you improve the precision of the TSSs call and align the nucleosome data to the center of divergent TSSs. This indicates that aligning to standard annotations can be misleading in assessing local chromatin architecture, and that promoters and enhancers share a similar nucleosome architecture.

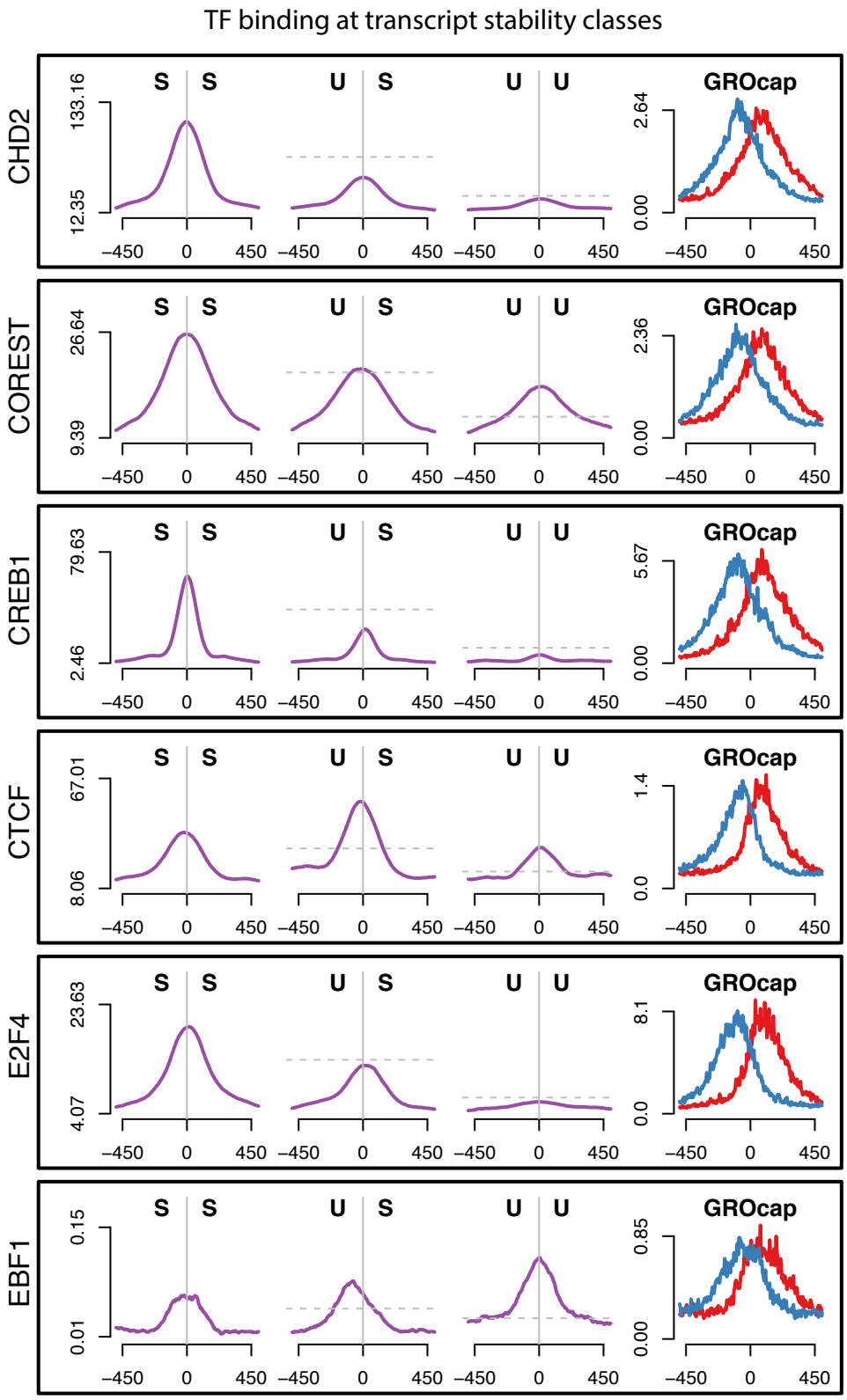
Core and Martins et al. Supplemental Figure 12 (part 1 of 14)



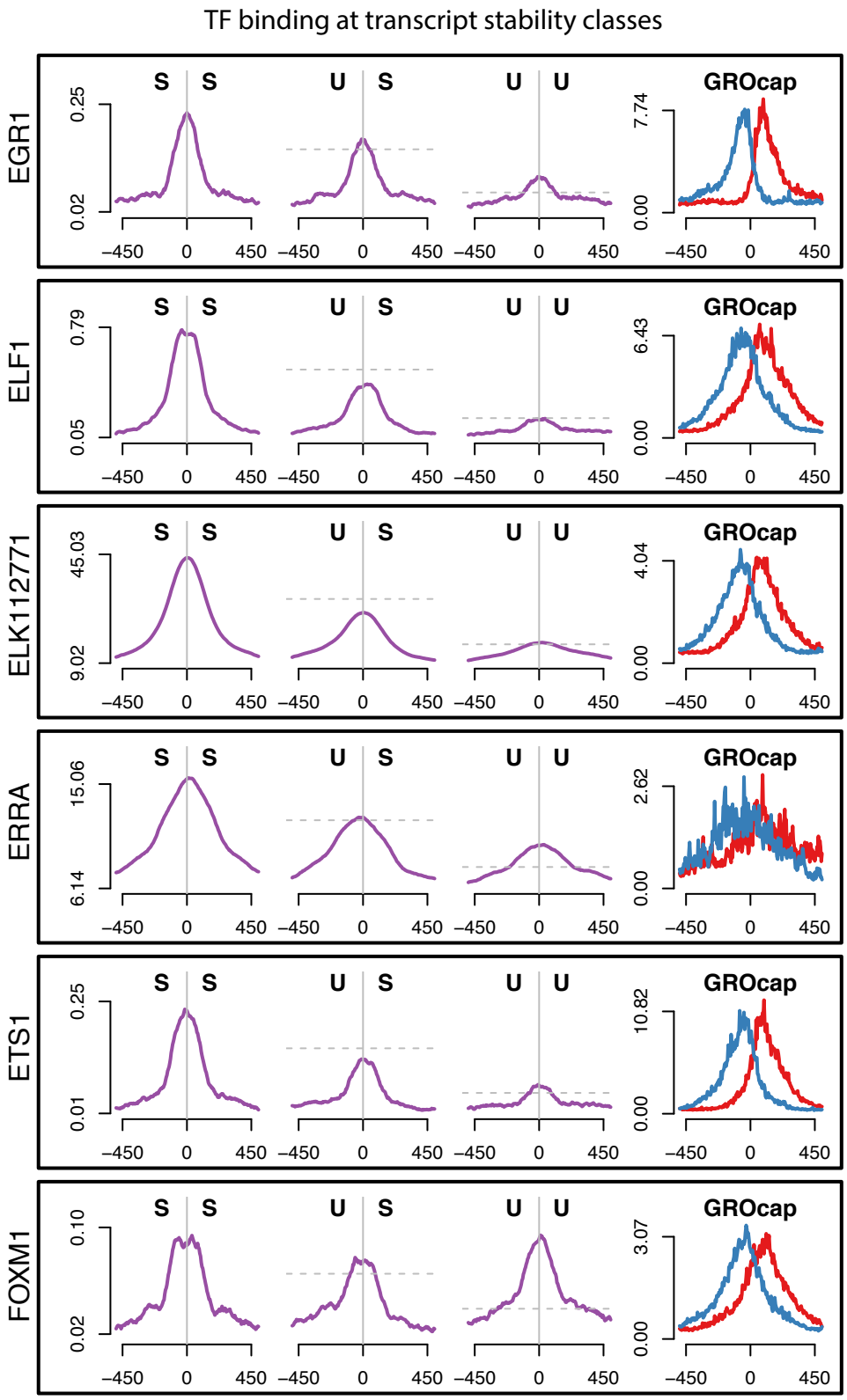
Core and Martins et al. Supplemental Figure 12 (part 2 of 14)



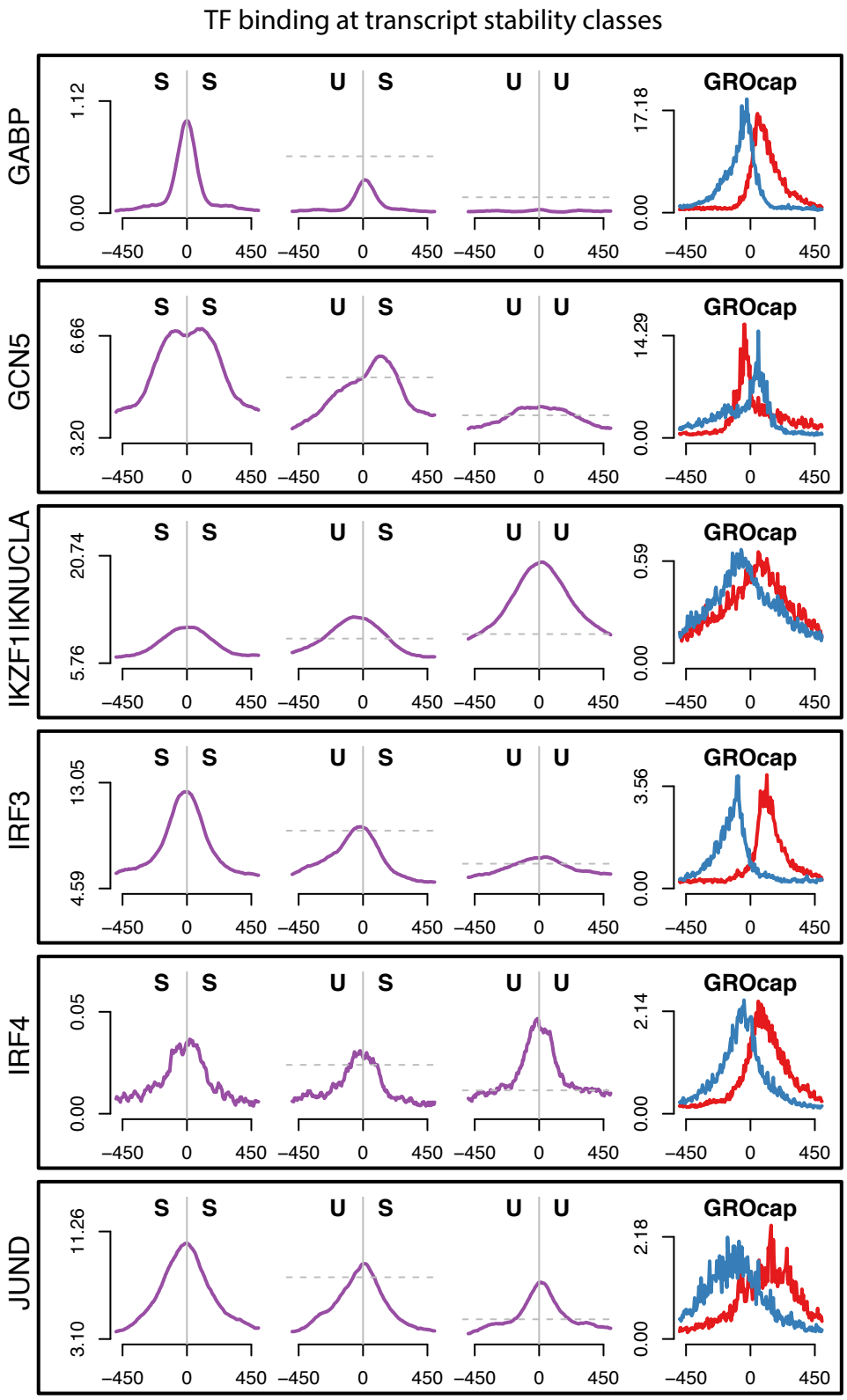
Core and Martins et al. Supplemental Figure 12 (part 3 of 14)



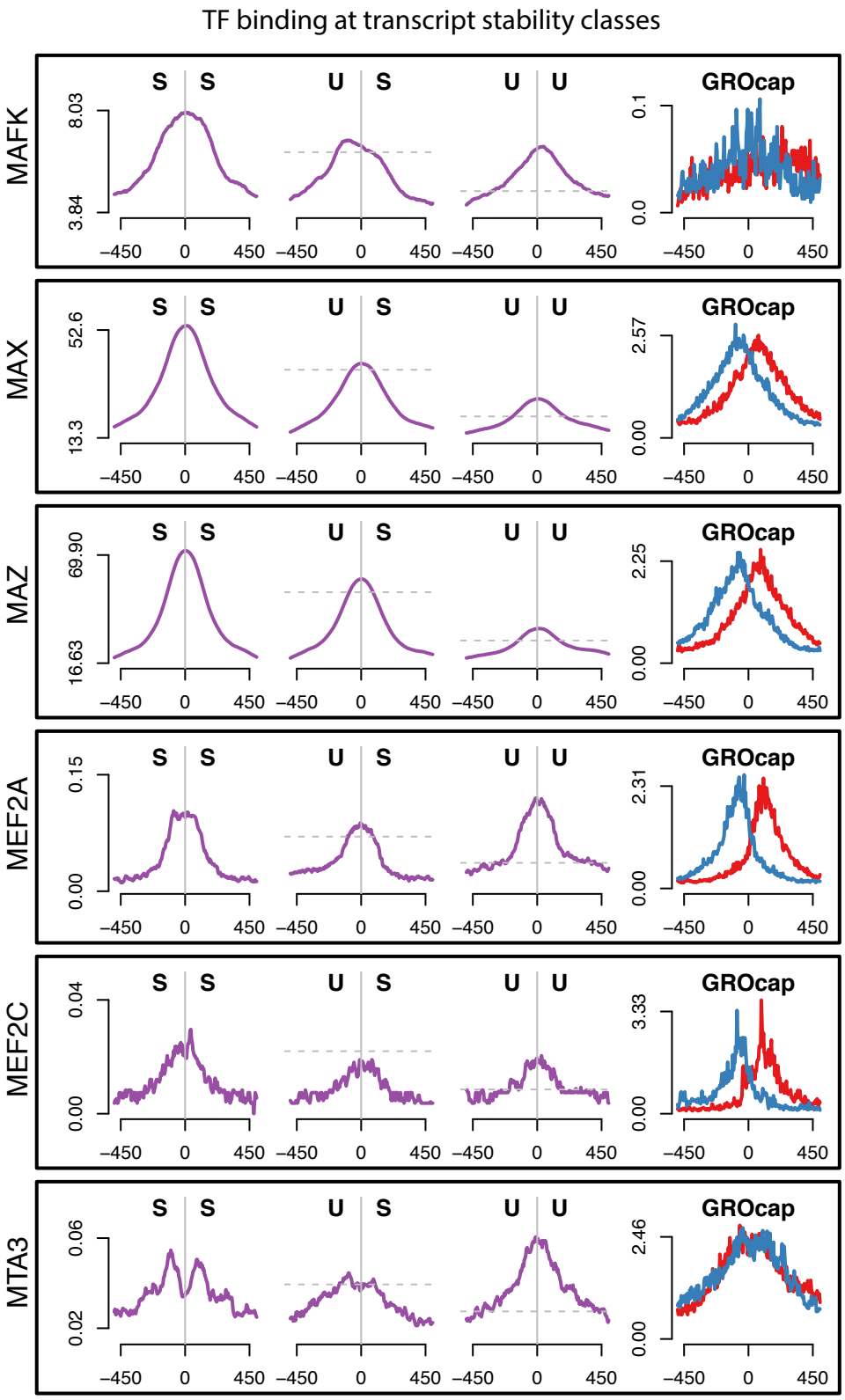
Core and Martins et al. Supplemental Figure 12 (part 4 of 14)



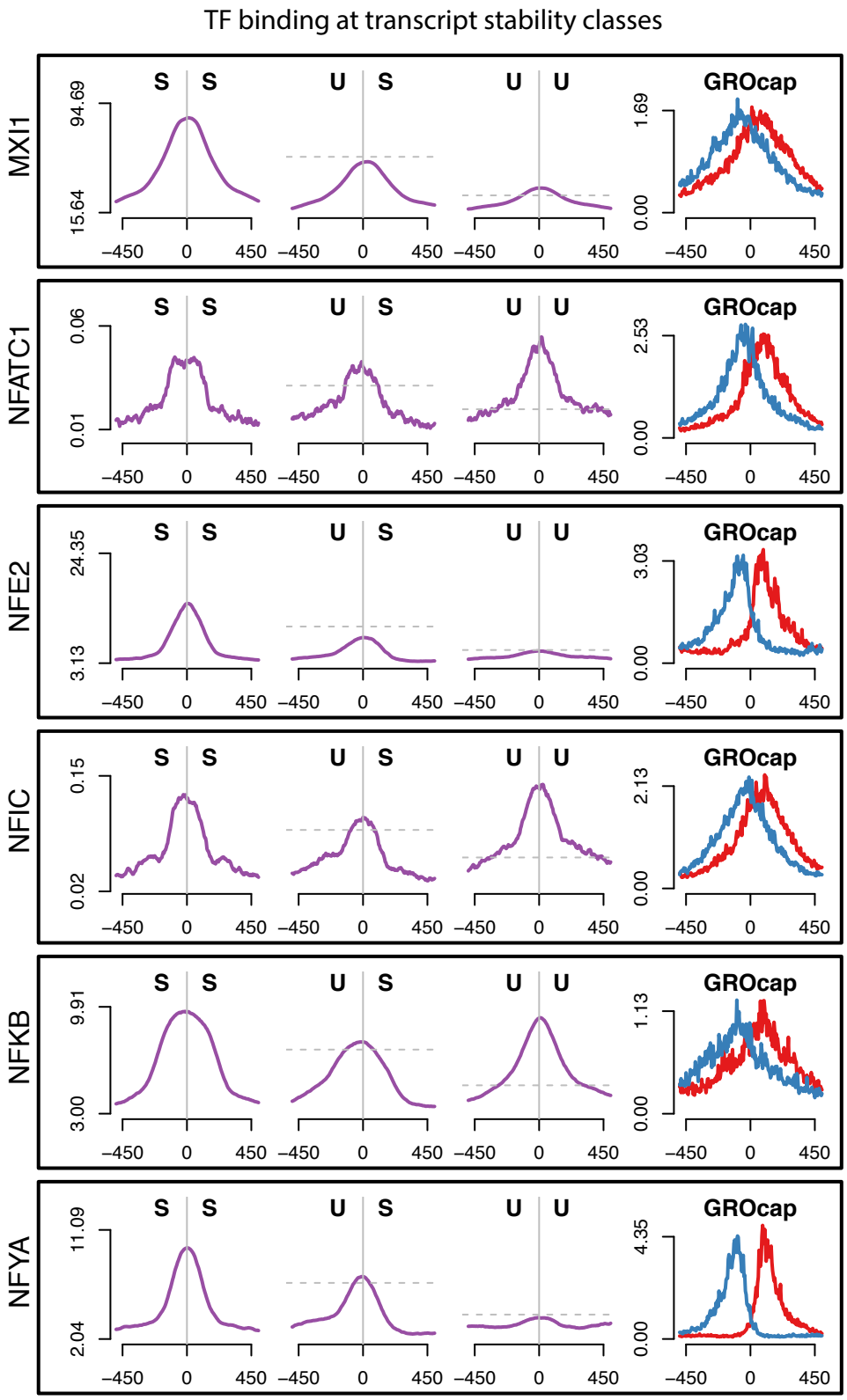
Core and Martins et al. Supplemental Figure 12 (part 5 of 14)



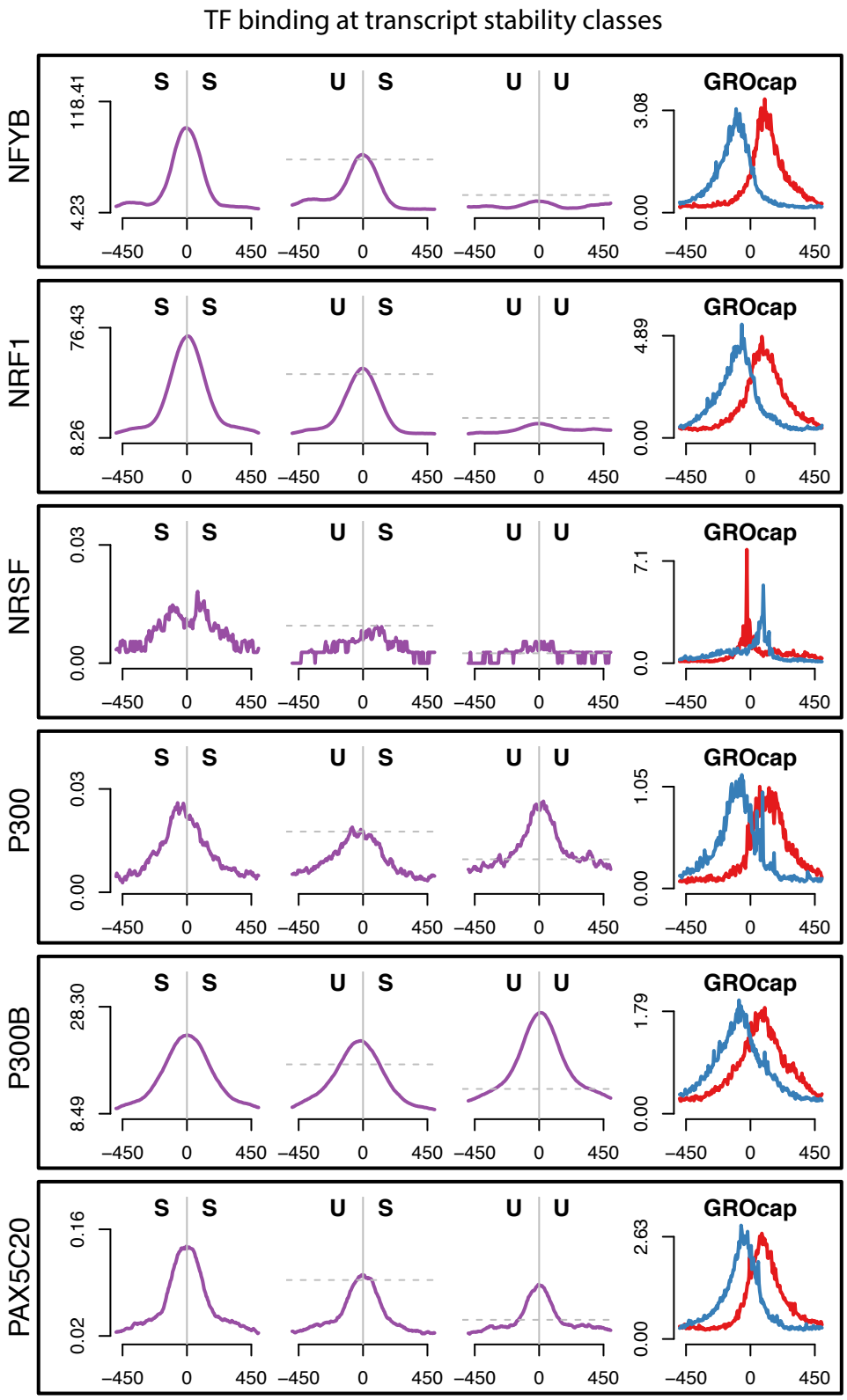
Core and Martins et al. Supplemental Figure 12 (part 6 of 14)



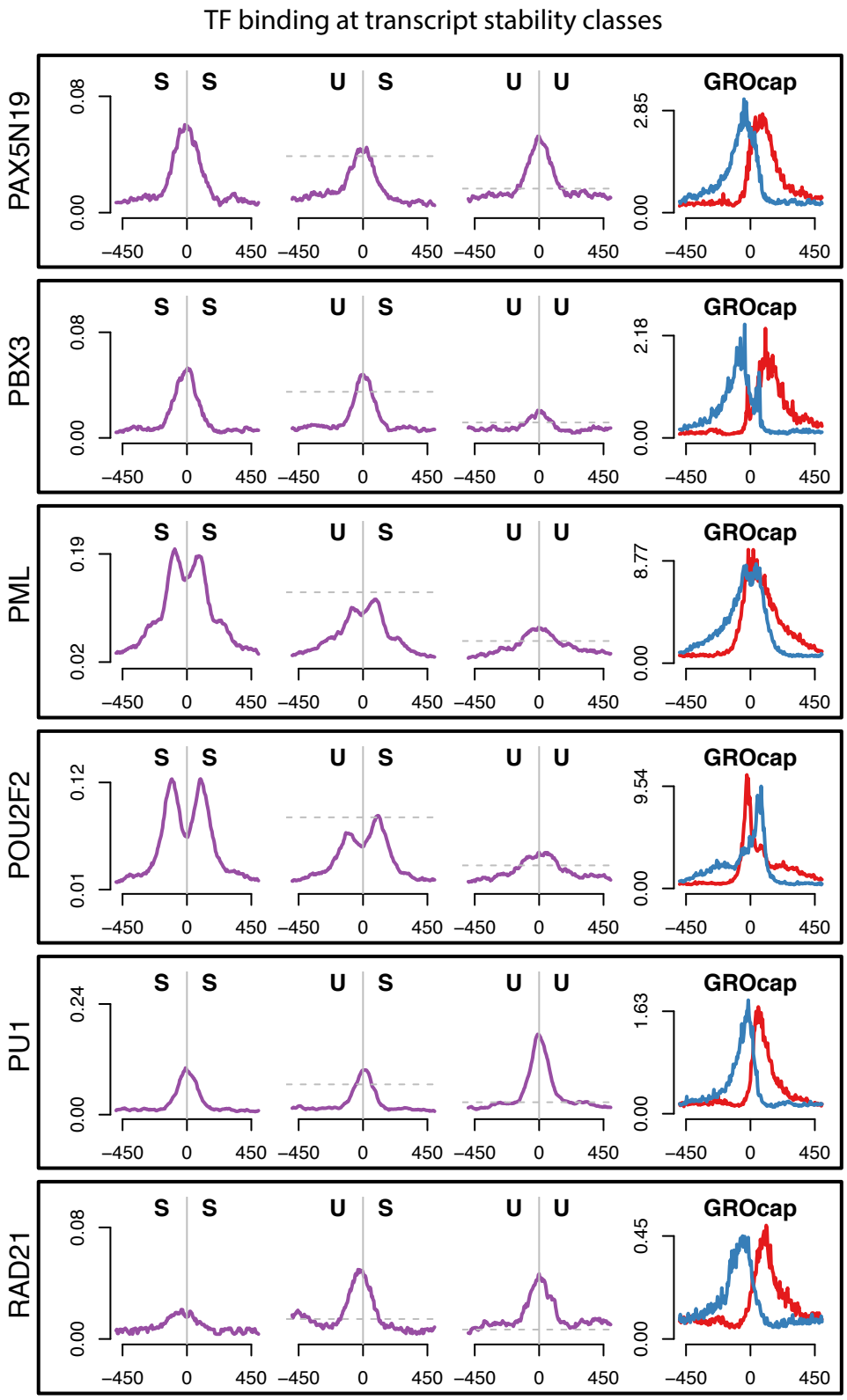
Core and Martins et al. Supplemental Figure 12 (part 7 of 14)



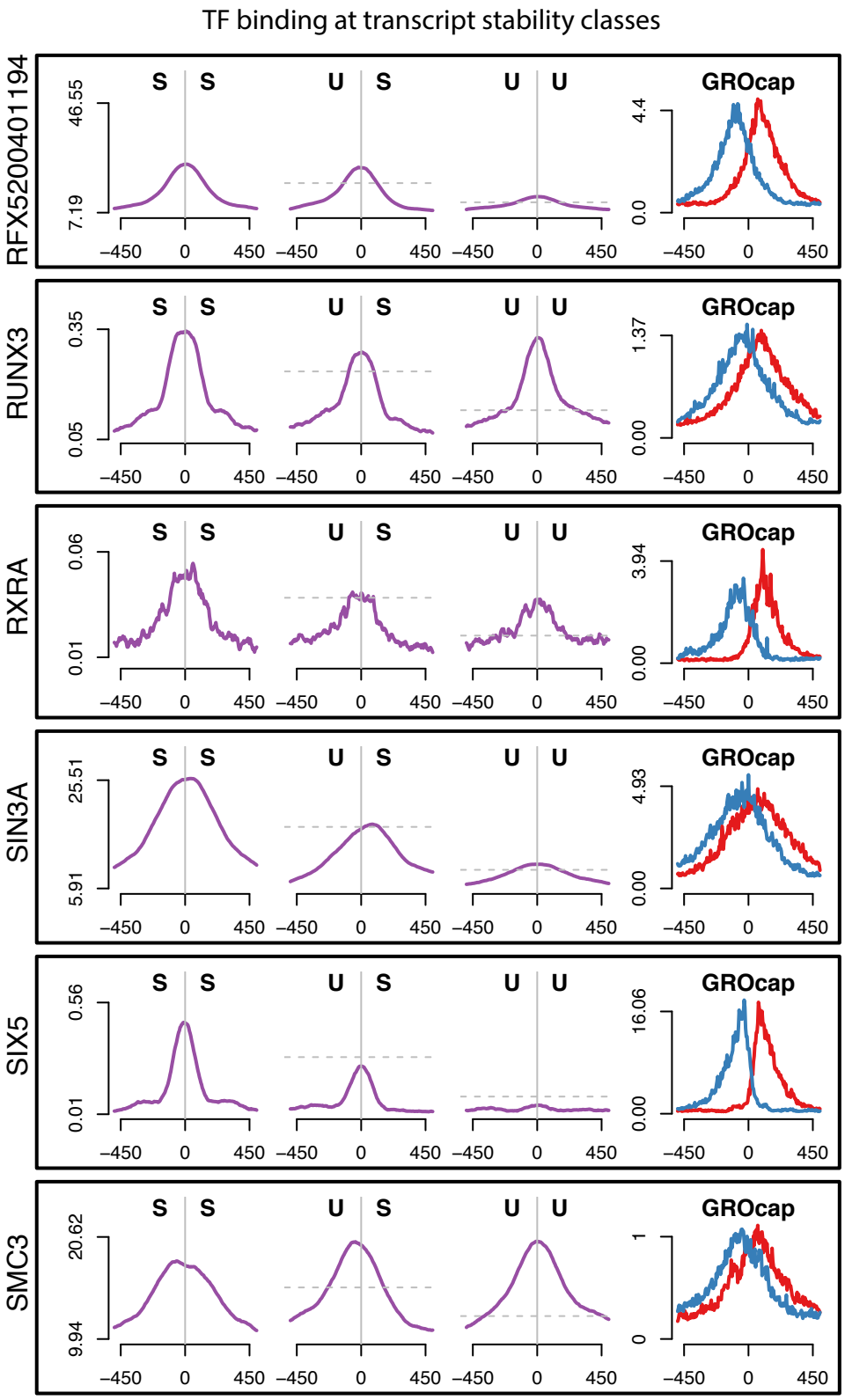
Core and Martins et al. Supplemental Figure 12 (part 8 of 14)



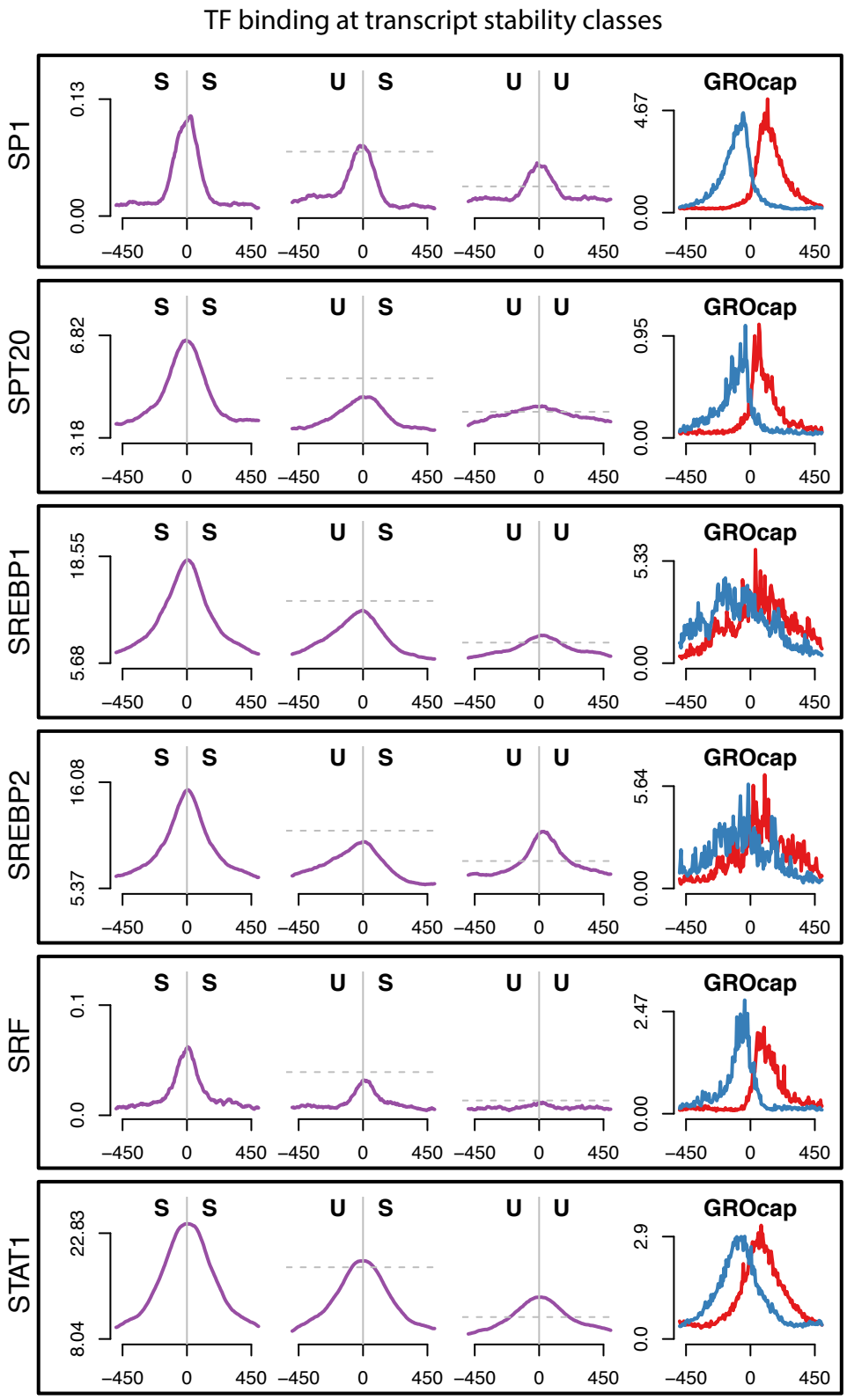
Core and Martins et al. Supplemental Figure 12 (part 9 of 14)



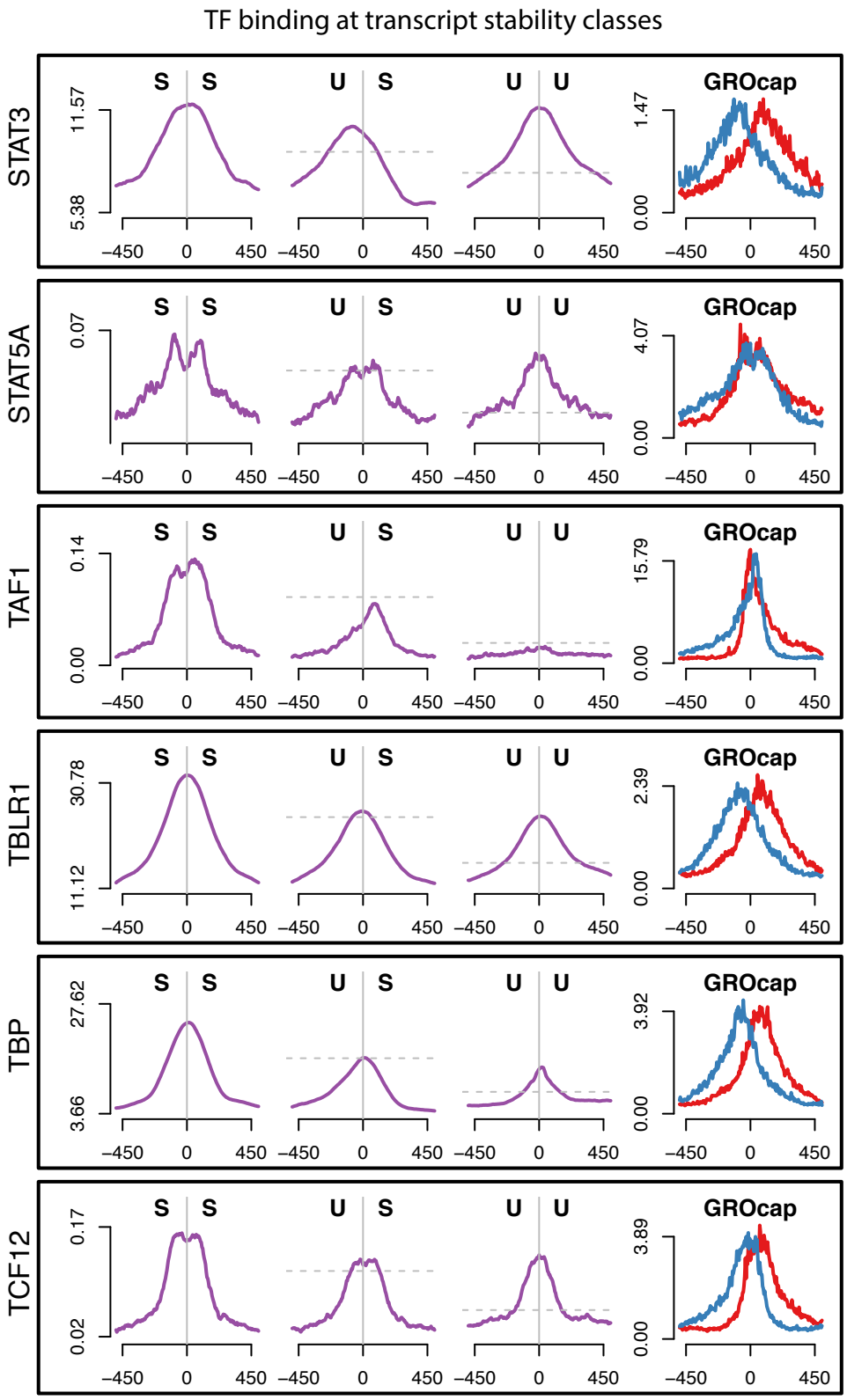
Core and Martins et al. Supplemental Figure 12 (part 10 of 14)



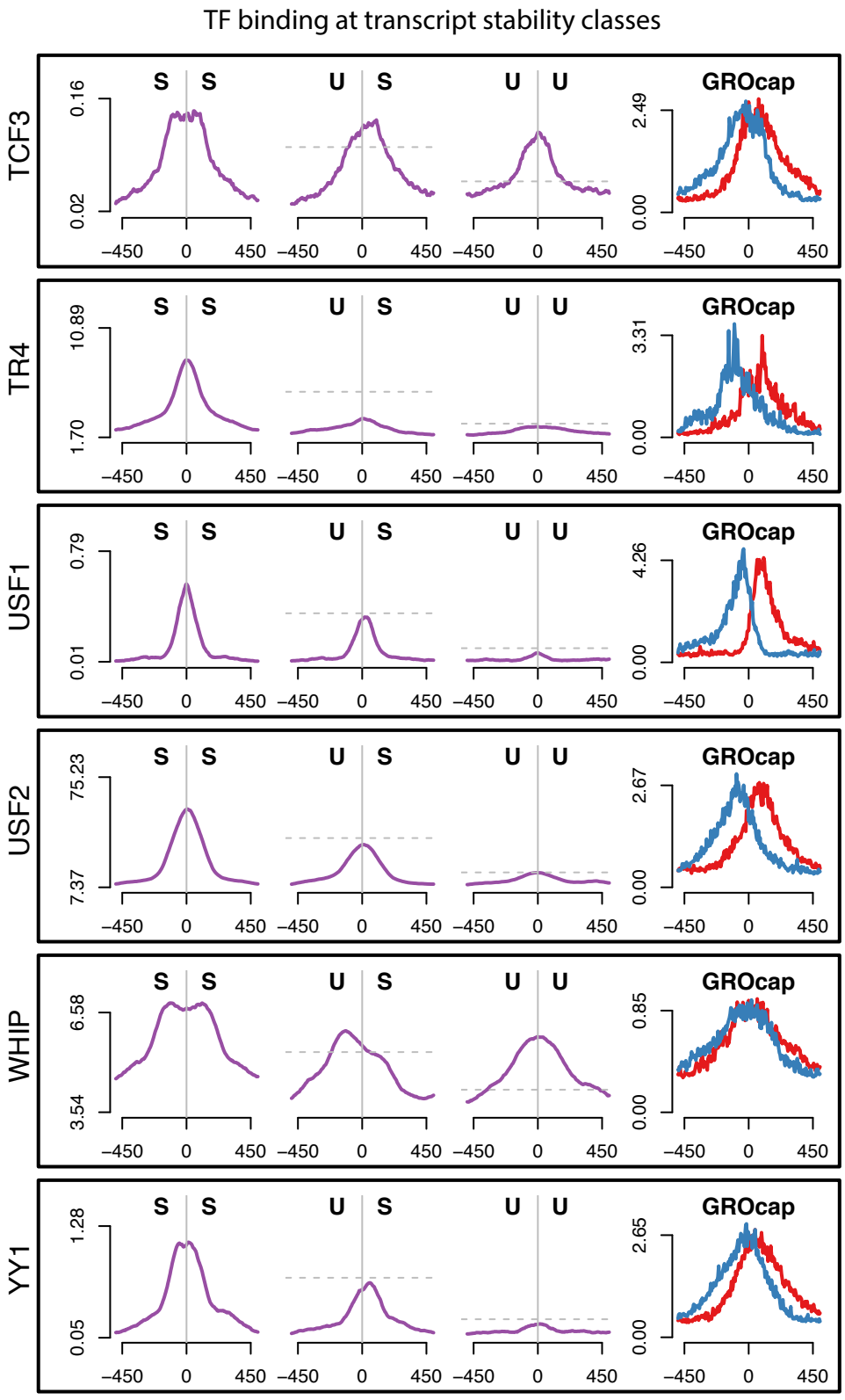
Core and Martins et al. Supplemental Figure 12 (part 11 of 14)

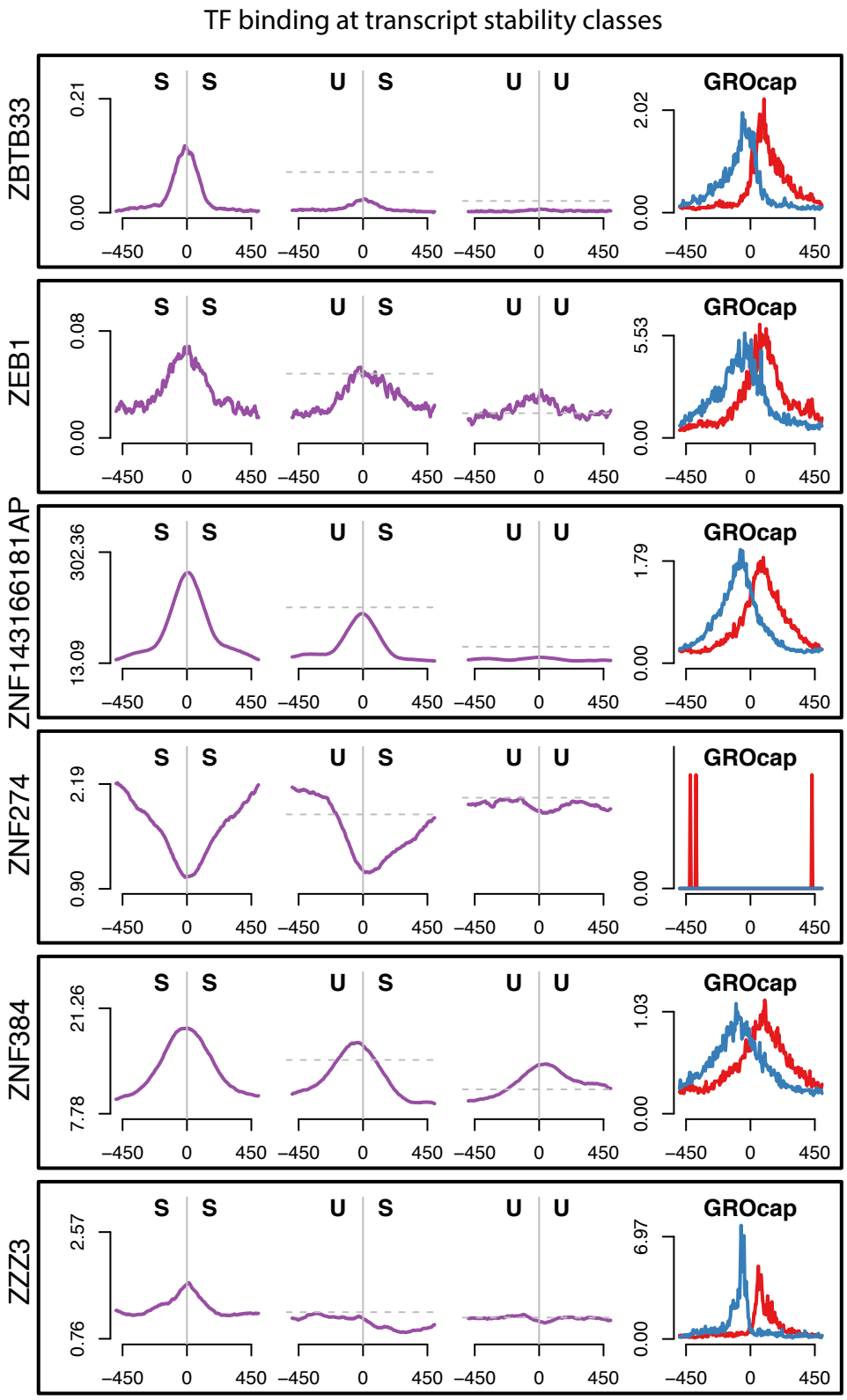


Core and Martins et al. Supplemental Figure 12 (part 12 of 14)



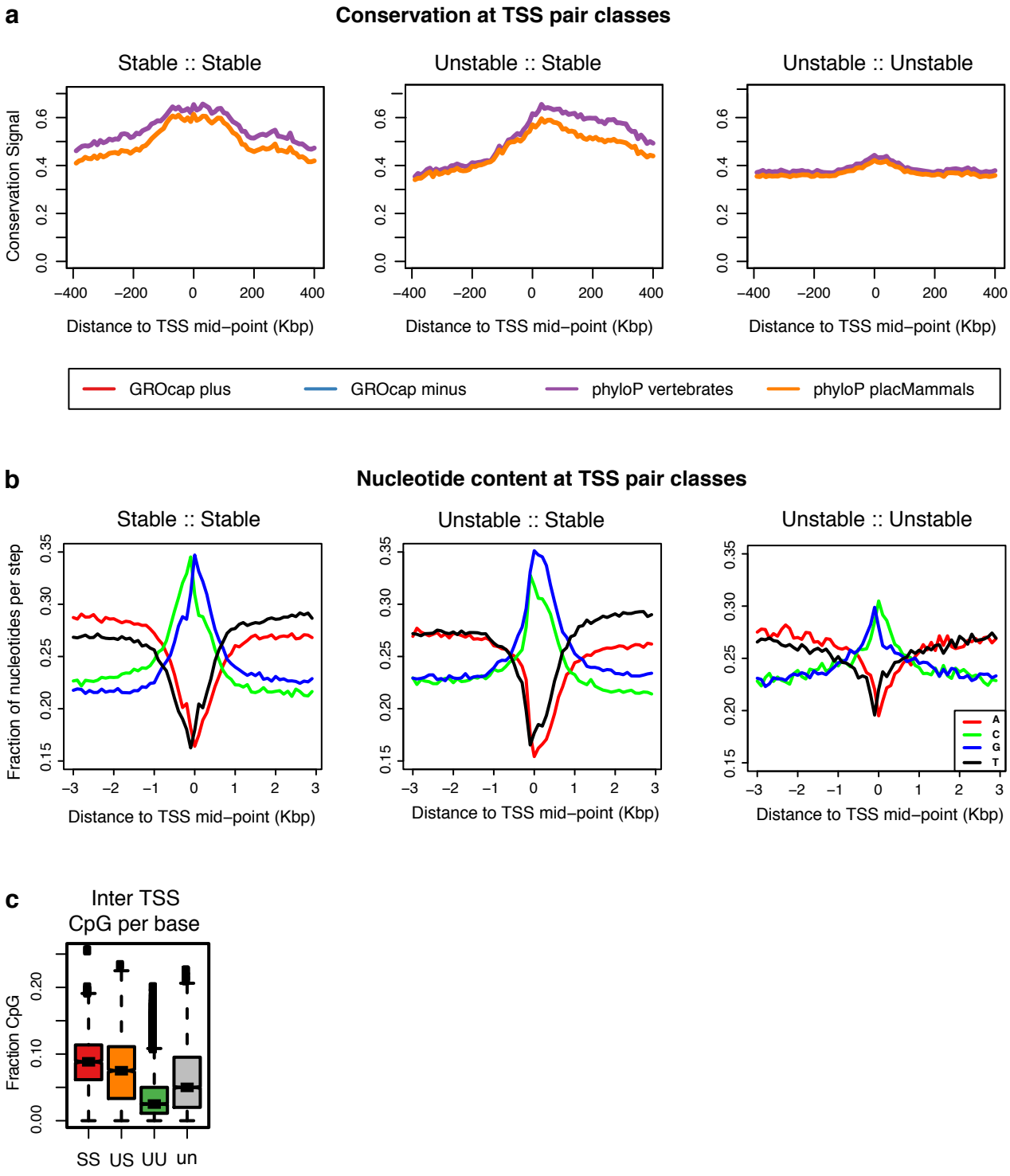
Core and Martins et al. Supplemental Figure 12 (part 13 of 14)





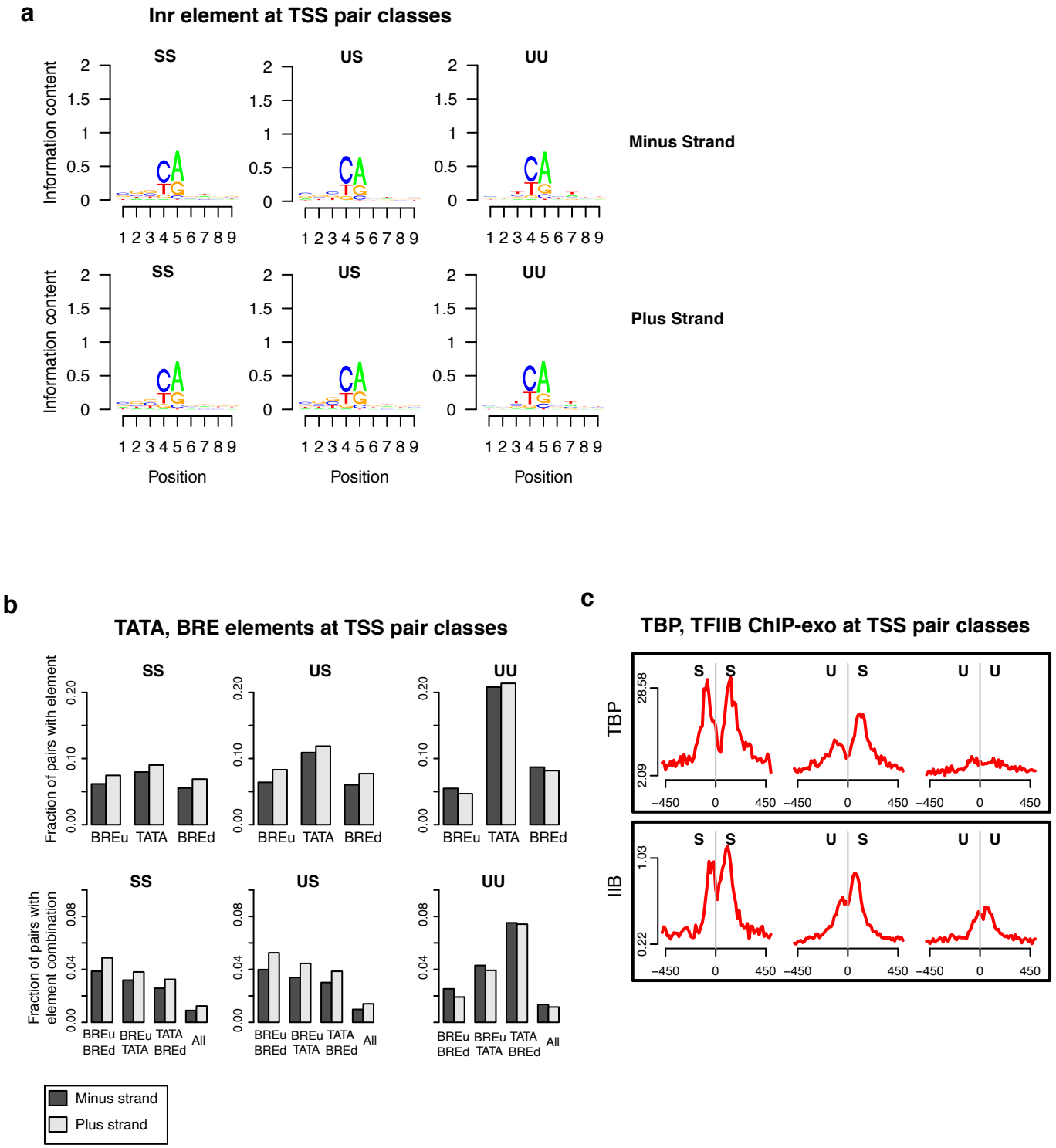
Supplemental Figure 12. Profiles of transcription factors at TSS pairs after stability classification
Composite profiles of ChIP-seq data for various transcription factors aligned to the center of GRO-cap TSS pairs after classifying pairs based on the stability of the transcript produced. Profiles are stable::stable, unstable::stable, unstable::unstable. The horizontal dashed lines represent the expected peak signal level if the signal followed the scaling of Pol II relative to the SS panel. The right panel shows GRO-cap data aligned to the peak of each individual transcription factor. All ChIP-seq data was produced by the ENCODE consortium in GM12878 cells.

Core and Martins et al. Supplemental Figure 13



Supplemental Figure 13. Sequence conservation and composition at paired TSS classes
(a) PhyloP (Pollard et al. Genome Res. 20, 110-121 (2010)) scores for vertebrates (purple) and placental mammals (orange), aligned to the center of GRO-cap TSS pairs after classifying pairs based on the stability of the transcript produced.
(b) Nucleotide frequencies aligned to the center of GRO-cap TSS pairs after classifying pairs based on the stability of the transcript produced. **(c)** Fraction of CpG dinucleotides in between divergent TSSs in pairs for different stability classes.

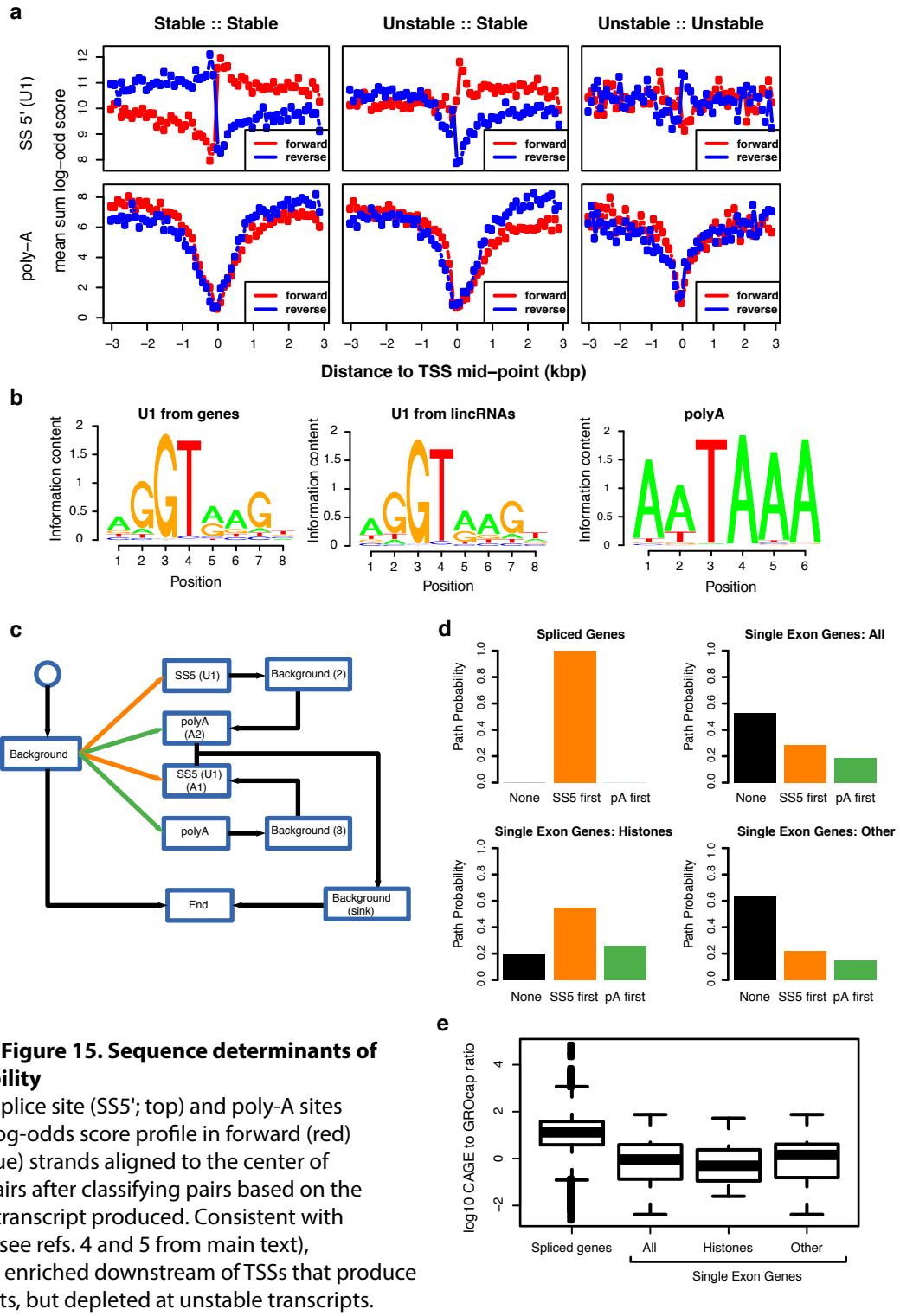
Core and Martins et al. Supplemental Figure 14



Supplemental Figure 14. Core promoter motifs and factors at TSSs

(a) Sequence logos showing INR element underlying both minus strand (top) and plus strand (bottom) TSSs at the different transcript stability classes. Logos obtained by alignment on base with strongest GRO-cap signal in each TSS region. **(b)** Occurrences of core promoter elements (TATA, BREd, BREu) at canonical positions. Top shows individual elements and bottom row shows combinations of elements. **(c)** ChIP-exo profiles for TBP and TFIIB (K562 cells) aligned to the center of GRO-cap TSS pairs after classifying pairs based on the stability of the transcript produced. Profiles are stable::stable (left), unstable::stable (center), unstable::unstable (right).

Core and Martins et al. Supplemental Figure 15



Supplemental Figure 15. Sequence determinants of transcript stability

(a) Five-prime splice site (SS5'; top) and poly-A sites (PAS; bottom) log-odds score profile in forward (red) and reverse (blue) strands aligned to the center of GRO-cap TSS pairs after classifying pairs based on the stability of the transcript produced. Consistent with previous work (see refs. 4 and 5 from main text), the SS5 motif is enriched downstream of TSSs that produce stable transcripts, but depleted at unstable transcripts. In contrast, the PAS motif is depleted downstream of stable TSSs. **(b)** PWM motifs for SS5' and PAS elements used in (a). SS5' PWMs obtained from GENCODE annotations with no apparent difference between protein-coding and lincRNAs. PAS PWM from Beaudoing et al. (see ref. 64 in methods). **(c)** HMM diagram for PAS versus SS5' relative motif position analysis. Boxes represent sequences of states representing the corresponding PWM motifs. Alternative paths capture the various possible relative element positions. **(d)** Estimated path posteriors through HMM for spliced gene transcripts and curated single exon gene transcripts. Single exon set is further split between histone coding transcripts and other. **(e)** GRO-cap to CAGE ratios in the subsets shown in (d).

Table S1: Summary of datasets and mapped reads generated for this study

Cell line	Assay	TAP used?	Length of mapped reads (bp)	# reads mapped
GM12878	GRO-cap	no TAP	30	6541296
GM12878	GRO-cap	with TAP	30	27314798
GM12878	GRO-seq	with TAP	30	105765321
K562	GRO-cap	with TAP	30	9267605
K562	GRO-cap	no TAP	30	26634162
K562	GRO-seq	with TAP	30	12721755
K562	PRO-seq	with TAP	15-100	364790421

Table S2: Classifications from the literature associated with TFs found in the 'TSS cluster'

factor	repressor	activator /Co-act	GTF
Chd1			
Gcn5		X	
Mta3	X		
Nrsf	X		
Pml	X		
Pou2f2	X		
Stat5a	X		
Taf1			X
Whip			
YY1	X	X	